

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 4, 2005, 20:11:25 ; Search time 3528.8 Seconds
(without alignments)
9927.764 Million cell updates/sec

Title: US-08-731-499-2
Perfect score: 723
Sequence: 1 TGAAGCTGTCATGGTTACC.....TGAATGGAATGATGGATTGCG 723

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl.*

- 1: gb_ba.*
- 2: gb_hgt.*
- 3: gb_in.*
- 4: gb_on.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	722.2	99.9	723	6 AR070325	Sequence 2 from patent US 5892010.
2	722.2	99.9	723	6 BD085726	Genes from the 20Q13 amplicon and their uses
3	714.8	98.9	163196	2 AC005220	Gray,J., Collins,C., Hwang,S.-i., Godfrey,T., Kowbel,D. and Rommens,J.
4	573.4	79.3	181313	2 AC021963	Genes from the 20Q13 amplicon and their uses
5	169	23.4	2808	6 C0726468	Patent: US 5892010-A 2 06-APR-1999;
6	167.6	23.2	349	6 BD060529	Location/Qualifiers
7	167.6	23.2	2813	6 C0776693	source
8	167.6	23.2	2813	6 AR371889	1. 723
9	167.6	23.2	2813	9 AF041260	/organism="unknown"
10	167.6	23.2	3066	6 BD085735	/mol_type="unassigned DNA"
11	137.8	19.1	168876	2 AC128535	
12	137.8	19.1	233233	2 AC094145	
13	127.4	17.6	199019	10 AL928812	
14	127.4	17.6	235411	2 AC084066	
15	86.6	12.0	2916	10 AY219233	
16	41.8	5.8	7218	6 I66494	
17	41.8	5.8	125020	9 AP429315	
18	41.4	5.7	5541	5 AB097138	
19	41.2	5.7	957	6 AX506150	

20	41.2	5.7	1284	8 ATH243888	Arabidops
21	41.2	5.7	1296	8 AB032865	Arabidops
22	41.2	5.7	1304	8 AJ010500	Arabidops
23	41.2	5.7	1356	8 AY084558	Arabidops
c 24	41.2	5.7	110000	2 AC112592	Rattus no
c 25	41.2	5.7	110000	2 AC112592_1	Continuation (2 of
26	40.6	5.6	104429	9 AP003787	Continuation (2 of
27	40.6	5.6	165330	10 AC124987	AP003787 Homo sapi
28	40.6	5.6	172880	9 AC027419	AC124987 Mus muscu
c 29	40.2	5.6	136867	9 AC108019	AC027419 Homo sapi
c 30	40	5.5	88931	8 AP006686	AC108019 Homo sapi
c 31	40	5.5	193870	10 AL731674	AP006686 Lotus cor
32	40	5.5	206467	2 AC133430	AL731674 Mouse DNA
33	40	5.5	219712	2 AC127196	AC133430 Rattus no
c 34	39.8	5.5	121688	9 AC092533	AC127196 Rattus no
c 35	39.8	5.5	150396	2 AC025489	AC092533 Homo sapi
c 36	39.8	5.5	158709	2 AC083827	AC025489 Homo sapi
c 37	39.8	5.5	164739	2 BX682546	AC083827 Homo sapi
c 38	39.6	5.5	110000	2 AC146444_3	BX682546 Homo sapi
c 39	39.6	5.5	167207	2 BX571817	Continuation (4 of
c 40	39.6	5.5	243657	2 AC124924	BX571817 Danio rer
c 41	39.6	5.5	264463	2 AC103132	AC124924 Rattus no
42	39.4	5.4	159254	2 AC084394	AC103132 Rattus no
43	39.4	5.4	175320	2 AC139292	AC084394 Homo sapi
44	39.4	5.4	209895	9 AC074391	AC139292 Mus muscu
c 45	39.4	5.4	211130	2 AC098891	AC074391 Homo sapi
					AC098891 Rattus no

ALIGNMENTS

RESULT 1	AR070325	723 bp	DNA	linear	PAT 18-FEB-2000
LOCUS	Sequence 2 from patent US 5892010.				
DEFINITION	AR070325				
ACCESSION	AR070325.1	GI:7221213			
VERSION	Unknown.				
KEYWORDS	Unclasseified.				
SOURCE	1 (bases 1 to 723)				
ORGANISM	Gray,J., Collins,C., Hwang,S.-i., Godfrey,T., Kowbel,D. and Rommens,J.				
REFERENCE	Genes from the 20Q13 amplicon and their uses				
AUTHORS	Patent: US 5892010-A 2 06-APR-1999;				
TITLE	Location/Qualifiers				
JOURNAL	source				
FEATURES	1. 723				
	/organism="unknown"				
	/mol_type="unassigned DNA"				

ORIGIN	Query Match	99.9%	Score 722.2;	DB 6;	Length 723;
	Best Local Similarity	100.0%;	Pred. No. 6.8e-192;		
	Matches	723;	Conservative	0;	Mismatches 0; Indels 0; Gaps 0;
Qy	1	TGGAAGCTGTCATGGTTACCGTCTCTAACCGTTGGACTCTTAAGAGAAATGATTATTCCTGG	60		
Db	1	TGGAAGCTGTCATGGTTACCGTCTCTAACCGTTGGACTCTTAAGAGAAATGATTATTCCTGG	60		
Qy	61	TTTCTAGACAGGCCAAATGTAATTCACCTACGTGGCAGATTAAGAGAGTGGGCTTACTAG	120		
Db	61	TTTCTAGACAGGCCAAATGTAATTCACCTACGTGGCAGATTAAGAGAGTGGGCTTACTAG	120		
Qy	121	ATTTGATTGGTATTGAGCATGCTCTGAATGACAGTCCCAAAAGGACCTCTTATCGT	180		
Db	121	ATTTGATTGGTATTGAGCATGCTCTGAATGACAGTCCCAAAAGGACCTCTTATCGT	180		
Qy	181	TCCTCCCTTTGGGGAAGGGCTTTTGGCAGTCCCAAAAGGACCTCTTATCGT	240		
Db	181	TCCTCCCTTTGGGGAAGGGCTTTTGGCAGTCCCAAAAGGACCTCTTATCGT	240		
Qy	241	ATTGGTCGCTGTGTACACAATAAGCATTTCTCCAGATGTGCTGTGTGTAAGAAATGGTC	300		

Db 241 ATTGGTGGCTTTACAACTAAGCATTAATCTCTCAAGATGTGCTGTGTGAATAATGGTC 300
QY 301 ATAGATTCAAACTGTAGCTACTATGTGACAGGGGGGAGCAAGGCCACCTTTGTAA 360
Db 301 ATAGATTCAAACTGTAGCTACTATGTGACAGGGGGGAGCAAGGCCACCTTTGTAA 360
QY 361 AACATGTTTGGGGGAATGTTTTGTTTTTCAATTTTCTTATTACCTGGCAAAATAATCCAG 420
Db 361 AACATGTTTGGGGGAATGTTTTGTTTTTCAATTTTCTTATTACCTGGCAAAATAATCCAG 420
QY 421 GTGGTGTGTAGTACCACTAGAGATTATAAGTCAAGGAAGTGAATCAGCCTTACAA 480
Db 421 GTGGTGTGTAGTACCACTAGAGATTATAAGTCAAGGAAGTGAATCAGCCTTACAA 480
QY 481 ACAGTGGACCTCAACGAAGGAGATGTCGACCTGAACCCACWGAAGCGAAACTCAAAAGA 540
Db 481 ACAGTGGACCTCAACGAAGGAGATGTCGACCTGAACCCACWGAAGCGAAACTCAAAAGA 540
QY 541 GAAGAAAGCAAAACCAAGAACCTCTCTGTATGRCGTTTCTCAGACAAATGTAAGCCCTTA 600
Db 541 GAAGAAAGCAAAACCAAGAACCTCTCTGTATGRCGTTTCTCAGACAAATGTAAGCCCTTA 600
QY 601 CTTCCAGTATAGAAACCTAAGATACCTAGACGGGCTTTTGGGAACAATGGGCTCATGCC 660
Db 601 CTTCCAGTATAGAAACCTAAGATACCTAGACGGGCTTTTGGGAACAATGGGCTCATGCC 660
QY 661 ACAGGTAGTAGGACATATTTGTAGTGTGTGTATGGAATGTAATGGAATATGGATT 720
Db 661 ACAGGTAGTAGGACATATTTGTAGTGTGTGTATGGAATGTAATGGAATATGGATT 720
QY 721 GCG 723
Db 721 GCG 723

RESULT 2
BD085726
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
PN JP 2001524802-A/2
PD 04-DEC-2001
PF 15-JUL-1997 JP 1998506264
PR 15-JUL-1996 US 08/680395,16-OCT-1996 US 08/731499 PR
PI JOE W GRAY, COLIN CONRAD COLLINS, SOO IN HWANG, TONY GODFREY, PI
DAVID KOWBEL,
PI JOHANNA ROMMENS
PC C12N15/11,C12Q1/68,A61K48/00
CC Description of Artificial Sequence:lb11
PH Key Location/Qualifiers
FT source 1..723
FT Location/Qualifiers
1..723
/organism="synthetic construct"
/mol_type="genomic DNA"
/db_xref="taxon:32630"

BD085726 723 bp DNA linear PAT 27-AUG-2002
Genes from the 20q13 amplicon and their uses.
BD085726
BD085726.1 GI:22631336
JP 2001524802-A/2.
synthetic construct
other sequences: artificial sequences.
Gray,J.W., Collins,C.C., Hwang,S.I., Godfrey,T., Kowbel,D. and Rommens,J.
Genes from the 20q13 amplicon and their uses
Patent: JP 2001524802-A 2 04-DEC-2001;
THE REGENTS OF THE UNIVERSITY OF CALIFORNIA
OS Artificial Sequence
PN JP 2001524802-A/2
PD 04-DEC-2001
PF 15-JUL-1997 JP 1998506264
PR 15-JUL-1996 US 08/680395,16-OCT-1996 US 08/731499 PR
PI JOE W GRAY, COLIN CONRAD COLLINS, SOO IN HWANG, TONY GODFREY, PI
DAVID KOWBEL,
PI JOHANNA ROMMENS
PC C12N15/11,C12Q1/68,A61K48/00
CC Description of Artificial Sequence:lb11
PH Key Location/Qualifiers
FT source 1..723
FT Location/Qualifiers
1..723
/organism="synthetic construct"
/mol_type="genomic DNA"
/db_xref="taxon:32630"

Query Match 99.9%; Score 722.2; DB 6; Length 723;
Best Local Similarity 100.0%; Pred. No. 6.8e-192;
Matches 723; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TGGAGCTGTATGTTTACCGTCTTAACGTTGGACTCTTAAGAAAATGATTATTCCTGG 60
Db 1 TGGAGCTGTATGTTTACCGTCTTAACGTTGGACTCTTAAGAAAATGATTATTCCTGG 60
QY 61 TTTCTAGACAGGCCAAATGTAATTCACCTACGTGGCAGATTAAAGAGGTGGGCTTACTAG 120
Db 61 TTTCTAGACAGGCCAAATGTAATTCACCTACGTGGCAGATTAAAGAGGTGGGCTTACTAG 120
QY 121 ATTTGATTTGGGATTTGAGCATGCTCTGAATGACAGTCCCCAAAAGAGCCTCTTATCCGT 180
Db 121 ATTTGATTTGGGATTTGAGCATGCTCTGAATGACAGTCCCCAAAAGAGCCTCTTATCCGT 180
QY 181 TCTTCCCTTTGGGNAAGGGCTTTTGGCACTTCCATGTCAATGTGGCAGTTGAGCTTGGAA 240
Db 181 TCTTCCCTTTGGGNAAGGGCTTTTGGCACTTCCATGTCAATGTGGCAGTTGAGCTTGGAA 240
QY 241 ATTGGTGTGTGTACACATTAAGCATTAATCTCTCAAGATGTGCTGTAGAAAATGGTC 300
Db 241 ATTGGTGTGTGTACACATTAAGCATTAATCTCTCAAGATGTGCTGTAGAAAATGGTC 300
QY 301 ATAGATTCAAACTGTAGCTACTATGTGACAGGGGGGAGCAAGGCCACCTTTGTAA 360
Db 301 ATAGATTCAAACTGTAGCTACTATGTGACAGGGGGGAGCAAGGCCACCTTTGTAA 360
QY 361 AACATGTTTGGGGGAATGTTTTGTTTTTCAATTTTCTTATTACCTGGCAAAATAATCCAG 420
Db 361 AACATGTTTGGGGGAATGTTTTGTTTTTCAATTTTCTTATTACCTGGCAAAATAATCCAG 420
QY 421 GTGGTGTGTAGTACCACTAGAGATTATAAGTCAAGGAAGTGAATCAGCCTTACAA 480
Db 421 GTGGTGTGTAGTACCACTAGAGATTATAAGTCAAGGAAGTGAATCAGCCTTACAA 480
QY 481 ACAGTGGACCTCAACGAAGGAGATGTCGACCTGAACCCACWGAAGCGAAACTCAAAAGA 540
Db 481 ACAGTGGACCTCAACGAAGGAGATGTCGACCTGAACCCACWGAAGCGAAACTCAAAAGA 540
QY 541 GAAGAAAGCAAAACCAAGAACCTCTCTGTATGRCGTTTCTCAGACAAATGTAAGCCCTTA 600
Db 541 GAAGAAAGCAAAACCAAGAACCTCTCTGTATGRCGTTTCTCAGACAAATGTAAGCCCTTA 600
QY 601 CTTCCAGTATAGAAACCTAAGATACCTAGACGGGCTTTTGGGAACAATGGGCTCATGCC 660
Db 601 CTTCCAGTATAGAAACCTAAGATACCTAGACGGGCTTTTGGGAACAATGGGCTCATGCC 660
QY 661 ACAGGTAGTAGGACATATTTGTAGTGTGTGTATGGAATGTAATGGAATATGGATT 720
Db 661 ACAGGTAGTAGGACATATTTGTAGTGTGTGTATGGAATGTAATGGAATATGGATT 720
QY 721 GCG 723
Db 721 GCG 723

RESULT 3
AC005220
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS

AC005220 163196 bp DNA linear PRI 01-JUL-1998
Homo sapiens chromosome 20, BAC clone 99 (LBNL H80), complete
sequence.
AC005220
AC005220.1 GI:3282159
HTG.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 163196)
Gray,J.W., Collins,C., Kimmerly,W., Bondoc,M., Cheng,J.,
Conolly,K.S., Gunning,K.M., Kadner,K., Miguel,T., Miller,C.,
Fitluck,S., Pollard,M., Rojeski,H., Subramanian,S. and Martin,C.H.

```

TITLE      Sequencing of human chromosome 20
JOURNAL    Unpublished
REFERENCE  2 (bases 1 to 163196)
AUTHORS    Rickes,D.O.
TITLE      Large Scale Sequence Analysis and Annotation with the Sequence
AUTHORS    Comparison Analysis (SCAN) System
JOURNAL    Unpublished
REFERENCE  3 (bases 1 to 163196)
AUTHORS    Gray,J.W., Collins,C., Kimmerly,W., Bondoc,M., Cheng,J.,
            Connolly,K.S., Gunning,K.M., Davis,C.A., Kadner,K., Miguel,T.,
            Piluck,S., Pollard,M., Rojeski,H., Subramanian,S. and Martin,C.H.
TITLE      Direct Submission
JOURNAL    Submitted (01-JUL-1998) Human Genome Center, DOE Joint Genome
            Institute, Lawrence Berkeley National Laboratory, MS 74-157,
            Berkeley, CA 94720, U.S.A.
COMMENT    Sequence submitted by:
FEATURES   Location/Qualifiers
            source
            1..163196
            /organism="Homo sapiens"
            /mol_type="genomic DNA"
            /db_xref="taxon:9606"
            /chromosome="20"
            /map="20"
            /clone="BAC 99"
            /note="IBNL H80"
            4588..4709
            /rpt_family="MIR"
            6685..6813
            /rpt_family="MIR"
            /rpt_family="trigger1"
            complement(6838..7220)
            /rpt_family="L1"
            7175..7314
            /rpt_family="L1"
            7322..7441
            /rpt_family="trigger1"
            complement(8695..9257)
            /rpt_family="MLT2B2"
            10121..10350
            /rpt_family="Alu"
            11057..11357
            /rpt_family="Alu"
            13431..13702
            /rpt_family="Alu"
            15911..15962
            /rpt_family="MIR"
            16029..16484
            /note="99% identity B54864 BAC end sequence"
            complement(16891..17025)
            /rpt_family="Alu"
            complement(17221..17682)
            /rpt_family="L1"
            complement(18500..18756)
            /rpt_family="L1"
            complement(18909..19161)
            /note="GRAIL 2 excellent exon, frame 1"
            19640..19943
            /rpt_family="Alu"
            complement(22025..22314)
            /rpt_family="Alu"
            complement(22903..23020)
            /note="GRAIL 2 excellent exon, frame 1"
            complement(23247..23567)
            /rpt_family="Alu"
            complement(24441..24514)
            /note="GRAIL 2 excellent exon, frame 0"
            25506..25525
            /note="(T)20"
            /rpt_type=tandem
            /rpt_unit="t"
            repeat_region
            complement(25538..25679)
            /rpt_family="MIR"
            complement(26920..27577)
            /rpt_family="L1"
            complement(28412..28563)
            /rpt_family="MER20"
            28734..29019
            /rpt_family="Alu"
            30266..30455
            /rpt_family="THE1"
            30789..30888
            /note="GRAIL 2 excellent exon, frame 2"
            31841..31972
            /note="GRAIL 2 excellent exon, frame 0"
            32270..32290
            /note="(T)21"
            /rpt_type=tandem
            /rpt_unit="t"
            32702..32860
            /note="GRAIL 2 excellent exon, frame 1"
            33268..33474
            /note="GRAIL 2 excellent exon, frame 0"
            33836..33863
            /note="(GT)14"
            /rpt_type=tandem
            /rpt_unit="gt"
            35158..35265
            /note="GRAIL 2 excellent exon, frame 0"
            complement(41855..42137)
            /rpt_family="Alu"
            43253..43465
            /note="GRAIL 2 excellent exon, frame 1"
            complement(50487..50600)
            /rpt_family="L1"
            50605..50855
            /rpt_family="L1"
            51461..51822
            /rpt_family="THE1"
            complement(54343..54856)
            /rpt_family="L1"
            complement(55490..55796)
            /rpt_family="Alu"
            55900..56027
            /rpt_family="MIR"
            complement(61007..61096)
            /note="GRAIL 2 excellent exon, frame 1"
            join(61724..61846,71254..71319,75055..75261)
            /standard_name="af74a01"
            /note="100% identity dbEST:AA644590"
            complement(62247..62746)
            /rpt_family="L1"
            complement(65258..65541)
            /rpt_family="Alu"
            65587..65882
            /rpt_family="Alu"
            68256..68282
            /note="(T)27"
            /rpt_type=tandem
            /rpt_unit="t"
            69077..69495
            /rpt_family="Alu"
            69537..69618
            /rpt_family="Alu"
            complement(70542..70721)
            /rpt_family="MERS"
            72072..72255
            /rpt_family="MIR"
            complement(72260..72340)
            /rpt_family="MLT1"
            74417..74456
            /note="(CT)20"
            /rpt_type=tandem
            /rpt_unit="ct"
            75057..75320
            /note="GRAIL 2 excellent exon, frame 2"
            join(75199..75323,84601..84746)

```

		/standard_name="zm96b07"	
		/note="96% identity dbEST:AA079436"	
repeat_region		complement(77059..77202)	
repeat_region		/rpt_family="MER5"	
repeat_region		complement(77269..77551)	
prim_transcript		/rpt_family="Alu"	
join(78685..79043,88203..88578)			
/standard_name="ym20b03"			
/note="97% identity dbEST:H15040"			
prim_transcript		79526..79590	
		/standard_name="ncI7d10"	
		/note="81% identity dbEST:AX259187"	
repeat_region		complement(81547..81827)	
repeat_region		/rpt_family="MIR"	
repeat_region		82700..82865	
repeat_region		/rpt_family="Alu"	
repeat_region		83191..83347	
repeat_region		/rpt_family="L1"	
repeat_region		complement(83468..83594)	
repeat_region		/rpt_family="MIR"	
repeat_region		84144..84436	
repeat_region		/rpt_family="Alu"	
prim_transcript		complement(84402..84774)	
		/standard_name="nII4b12"	
prim_transcript		/note="98% identity dbEST:AAS54549"	
prim_transcript		complement(85065..85162)	
		/standard_name="yw21h10"	
prim_transcript		/note="90% identity dbEST:H88501"	
prim_transcript		join(85165..85589,86249..86579,87354..87476)	
		/standard_name="yg09d02"	
misc_feature		/note="97% identity dbEST:R43020"	
misc_feature		complement(86451..86577)	
prim_transcript		/notes="GRAIL 2 excellent exon, frame 2"	
prim_transcript		complement(87953..88445)	
prim_transcript		/standard_name="zu90c12"	
misc_feature		/note="99% identity dbEST:AA626177"	
misc_feature		complement(89695..89866)	
repeat_region		/notes="GRAIL 2 excellent exon, frame 1"	
repeat_region		89959..90146	
repeat_region		/rpt_family="MER5"	
Query Match			
Best Local Similarity 98.9%; Score 714.8; DB 9; Length 163196;			
Matches 716; Conservative 2; Mismatches 4; Indels 0; Gaps 0;			
Qy	1	TGGAAGCTGTATGGCGTTCACGTTCCACTCCTTAAGAAAATGATTATTCCTGC 60	
Db	61259	TGGAAGCTGTATGGCGTTCACGTTCCACTCCTTAAGAAAATGATTATTCCTGC 61318	
Qy	61	TTTTTAGACAGGCCAATGTAATTCACCTACGTGGCAGATTAAAGNGTGCGCTTACTAG 120	
Db	61319	TTTTTAGACAGGCCAATGTAATTCACCTACGTGGCAGATTAAAGNGTGCGCTTACTAG 61378	
Qy	121	ATTTGATTGGGTATTGAGCATGCTCTCAATGACAGTCCCACAAAAAGGACCCTCTATCCGT 180	
Db	61379	ATTTGATTGGGTATTGAGCATGCTCTCAATGACAGTCCCACAAAAAGGACCCTCTATCTGT 61438	
Qy	181	TCTTCCCCCTGGGGAAAGGGCTTTTGGCACCTTCCATGTCGAATGTGGCAGTTGAGCTTGAA 240	
Db	61439	TCTTCCCCCTGGGGAAAGGGATTTTGGCACCTTCCATGTCGAATGTGGCAGTTGAGCTTGAA 61498	
Qy	241	ATTGGTGC GTTGACACATAAGCATTACTTCCAAAGATGTCCTGTGTAGAATGTC 300	
Db	61499	ATTGGTGC GTTGACACATAAGCATTACTTCCAAAGATGTCCTGTGTAGAATGTC 61558	
Qy	301	ATAGATTCAAACCTGTAGCTACTATGTGACAGGGGGGCGAACAGGCCACCCACTTTGTAA 360	
Db	61559	ATAGATTCAAACCTGTAGCTACTATGTGACAGGGGGGCGAACAGGCCACCCACTTTGTAA 61618	
Qy	361	AACATGTTTTGGGGGAATGTTTTGTTTTTCACTTTCTATTAATCTGCGOAAAATAATCCAG 420	
Db	61619	AACATGTTTTGGGGGAATGTTTTGTTTTTCACTTTCTATTAATCTGCGOAAAATAATCCAG 61678	

Qy	421	GTGGTGTGTGAGTCCACCTAGTAGAGATTTAAAGTCCAAAGGAAGTAGAATCAGCCCTTACAA	480
Db	61679	GTGGTGTGTGAGTCCACCTAGTAGAGATTTAAAGTCCAAAGGAAGTAGAATCAGCCCTTACAA	61738
Qy	481	ACAGTGGACCTCAACGAGGAGATCTGCACCTCTGATGRCGTTTCTCAGACAAATGGTAAGCCCTTAA	540
Db	61739	ACAGTGGACCTCAACGAGGAGATCTGCACCTGAAACCCACAGAAAGCAAACTCAAAAGA	61798
Qy	541	GAAGAAAGCAAAACCAAGAACCTCTCTGATGRCGTTTCTCAGACAAATGGTAAGCCCTTAA	600
Db	61799	GAAGAAAGCAAAACCAAGAACCTCTCTGATGRCGTTTCTCAGACAAATGGTAAGCCCTTAA	61858
Qy	601	CTTCCAGTATAGAAAACCTAAGATACCTAGACGGCGCTTTTGGGAAACAATGGGCTCATGCC	660
Db	61859	CTTCCAGTATAGAAAACCTAAGATACCTAGACGGCGCTTTTGGGAAACAATGGGCTCATGCC	61918
Qy	661	ACAGGTAGTAGGAGACATAAATCTGACGTGCTGTGTATGGAATGTCAATGGATATGGATT	720
Db	61919	ACAGGTAGTAGGAGACATAAATCTGACGTGCTGTGTATGGAATGTCAATGGATATGGATT	61978
Qy	721	GC 722	
Db	61979	GC 61980	
RESULT 4			
AC021963/c			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
COMMENT			

QY 481 ACAGTGGACCTCAACGAAGGAGATGCTGCACCTGAACCCACGAAAGCGAAACTCAAAAGA 540
|||||
Db 37065 ACAGTGGACCTCAACGAAGGAGATGCTGCACCTGAACCCACGAAAGCGAAACTCAAAAGA 37006
|||||
QY 541 GAAGAAGCAAAACCAAGAACTCTCTGATGRCGTTTCTC 579
|||||
Db 37005 GAAGAAGCAAAACCAAGAACTCTCTGATGGCGTTTCTC 36967
|||||
RESULT 5
CQ726468 2808 bp DNA linear PAT 03-FEB-2004
LOCUS Sequence 12402 from Patent WO02068579.
DEFINITION CQ726468
ACCESSION CQ726468
VERSION CQ726468.1 GI:42289602
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Venter,C.J., Adams,M.C., Li,P.W. and Myers,E.W.
TITLE Kits, such as nucleic acid arrays, comprising a majority of
humanexons or transcripts, for detecting expression and other uses
thereof
JOURNAL Patent: WO 02068579-A 12402 06-SEP-2002;
PE Corporation (NY) (US)
FEATURES
source Location/Qualifiers
1..2808
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Query Match 23.4%; Score 169; DB 6; Length 2808;
Best Local Similarity 95.0%; Pred. No. 2.6e-36;
Matches 172; Conservative 2; Mismatches 7; Indels 0; Gaps 0;
QY 408 CAAATAATCCAGGTGGTGTGAGTCACCAAGTAGAGATTATAAGTCCCAAGGAAGTAGA 467
Db 1279 CCAAAAGGACCTGTGGTGTGAGTCACCAAGTAGAGATTATAAGTCCCAAGGAAGTAGA 1338
|||||
QY 468 ATCAGGCTTACAAACAGTGACCTCAACGAAGGAGATGCTGCACCTGAAACCCACGAAAGC 527
|||||
Db 1339 ATCAGGCTTACAAACAGTGACCTCAACGAAGGAGATGCTGCACCTGAAACCCACGAAAGC 1398
|||||
QY 528 GAAACTCAAAAGAGAGAAAGCAAAACCAAGAACTCTCTGATGRCGTTTCTCAGACAAAT 587
Db 1399 GAAACTCAAAAGAGAGAAAGCAAAACCAAGAACTCTCTGATGRCGTTTCTCAGACAAAT 1458
|||||
QY 588 G 588
Db 1459 G 1459
RESULT 6
BD060529 349 bp DNA linear PAT 27-AUG-2002
LOCUS Secreted expressed sequence tags (seSTs).
DEFINITION BD060529
ACCESSION BD060529
VERSION BD060529.1 GI:22606135
KEYWORDS Zea mays
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 349)
AUTHORS Jacobs,K., Mccoy,J.M., Lavallie,E.R., Racie,L.A., Merberg,D.,
Treacy,M., Spaulding,V. and Agostino,M.J.
TITLE Secreted expressed sequence tags (seSTs)
JOURNAL Patent: JP 2001518793-A 889 16-OCT-2001;

GENETICS INSTITUTE INC
PN JP 2001518793-A/889
PD 16-OCT-2001
PF 10-APR-1998 JP 1998543070
PR 10-APR-1997 US 08/837312
PI KENNETH JACOBS, JOHN M MCCOY, EDWARD R LAVALLIE, LISA A RACIE, PI
DAVID MERBERG,
PI MAURICE TREACY, VIKKI SPAULDING, MICHAEL J AGOSTINO PC
C12N15/12, C12N5/10, C07K14/47, C12Q1/68, A61K38/17 CC Strandedness:
Double;
CC Topology: Linear;
FH Key Location/Qualifiers.
FEATURES
source Location/Qualifiers
1..349
/organism="Zea mays"
/mol_type="genomic DNA"
/db_xref="taxon:4577"
ORIGIN
Query Match 23.2%; Score 167.6; DB 6; Length 349;
Best Local Similarity 98.2%; Pred. No. 7.1e-36;
Matches 167; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 419 AGTGGTGTGAGTCACCAAGGAGATGCTGCACCTGAAACCCACGAAAGCGAAACTCAAAA 478
Db 69 ATGTGGTGTGAGTCACCAAGGAGATGCTGCACCTGAAACCCACGAAAGCGAAACTCAAAA 128
|||||
QY 479 AAACAGTGGACCTCAACGAAGGAGATGCTGCACCTGAAACCCACGAAAGCGAAACTCAAAA 538
Db 129 AAACAGTGGACCTCAACGAAGGAGATGCTGCACCTGAAACCCACGAAAGCGAAACTCAAAA 188
|||||
QY 539 GAGAAGAAAGCAAAACCAAGAACTCTCTGATGRCGTTTCTCAGACAAATG 588
Db 189 GAGAAGAAAGCAAAACCAAGAACTCTCTGATGRCGTTTCTCAGACAAATG 238
|||||
RESULT 7
CQ776693 2813 bp DNA linear PAT 11-MAR-2004
LOCUS Sequence 379 from Patent EPI394274.
DEFINITION CQ776693
ACCESSION CQ776693
VERSION CQ776693.1 GI:45380083
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Ohtani,N., Sugita,Y., Yamaya,M., Kubo,H., Nagai,H. and Izuwara,K.
TITLE Methods of testing for bronchial asthma or chronic obstructive
pulmonary disease
JOURNAL Patent: EP 1394274-A 379 03-MAR-2004;
Genox Research, Inc. (JP)
FEATURES
source Location/Qualifiers
1..2813
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Query Match 23.2%; Score 167.6; DB 6; Length 2813;
Best Local Similarity 98.2%; Pred. No. 6.4e-36;
Matches 167; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 419 AGTGGTGTGAGTCACCAAGGAGATGCTGCACCTGAAACCCACGAAAGCGAAACTCAAAA 478
Db 1298 ATGTGGTGTGAGTCACCAAGGAGATGCTGCACCTGAAACCCACGAAAGCGAAACTCAAAA 1357
|||||
QY 479 AAACAGTGGACCTCAACGAAGGAGATGCTGCACCTGAAACCCACGAAAGCGAAACTCAAAA 538
Db 1358 AAACAGTGGACCTCAACGAAGGAGATGCTGCACCTGAAACCCACGAAAGCGAAACTCAAAA 1417
|||||
QY 539 GAGAAGAAAGCAAAACCAAGAACTCTCTGATGRCGTTTCTCAGACAAATG 588
|||||

```
Db 1418 GAGAGAAAGCAACCAAGAACCTCTCTGATGGCGTTTCTCAGACAAATG 1467
|||||
RESULT 8
AR371889 2813 bp DNA linear PAT 12-SEP-2003
LOCUS
DEFINITION Sequence 3 from patent US 6395544.
ACCESSION AR371889
VERSION AR371889.1 GI:34608999
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 2813)
AUTHORS Cowser, L.M. and Freier, S.M.
TITLE Antisense modulation of BCAS1 expression
JOURNAL Patent: US 6395544-A 3 28-MAY-2002;
FEATURES
source Location/Qualifiers
1..2813
/organism="unknown"
/mol_type="genomic DNA"
ORIGIN
Query Match 23.2%; Score 167.6; DB 6; Length 2813;
Best Local Similarity 98.2%; Pred. No. 6.4e-36;
Matches 167; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Qy 419 AGGTGGTGTGTGAGTCACAGGAGATTTAAAGTCCAGGAAGTAGAATCAGCCTTAC 478
Db 1298 ATGTGGTGTGTGAGTCACAGGAGATTTAAAGTCCAGGAAGTAGAATCAGCCTTAC 1357
Qy 479 AAACAGTGGACCTCAACGAAGGAGATGCTGCACCTGAAACCCACGAAAGCGAAACTCAAAA 538
Db 1358 AAACAGTGGACCTCAACGAAGGAGATGCTGCACCTGAAACCCACGAAAGCGAAACTCAAAA 1417
Qy 539 GAGAGAAAGCAACCAAGAACCTCTCTGATGRCGTTTCTCAGACAAATG 588
Db 1418 GAGAGAAAGCAACCAAGAACCTCTCTGATGRCGTTTCTCAGACAAATG 1467
|||||
RESULT 9
AF041260 2813 bp mRNA linear PRI 22-JUL-1998
LOCUS
DEFINITION Homo sapiens AIB1 (AIB1) mRNA, complete cds.
ACCESSION AF041260
VERSION AF041260.1 GI:3335398
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2813)
AUTHORS Collins, C., Rommens, J.M., Kowbel, D., Godfrey, T., Tanner, M.,
Hwang, S.-I., Polikoff, D., Nonet, G., Cochran, J., Myambo, K.,
Jay, K.E., Froula, J., Cloutier, T., Kuo, W.-L., Yaswen, P., Dairkee, S.,
Giovannola, J., Hutchinson, G.B., Isola, J., Kallioniemi, O.-P.,
Palazzolo, M., Martin, C., Ericsson, C., Pinkel, D., Albertson, D.,
Li, W.-B. and Gray, J.W.
TITLE Positional cloning of ZNF217 and NAB1: genes amplified at 20q13.2
and overexpressed in breast carcinoma
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 95 (15), 8703-8708 (1998)
MEDLINE 98337979
PUBMED 9671742
REFERENCE 2 (bases 1 to 2813)
AUTHORS Collins, C., Rommens, J.M., Kowbel, D., Godfrey, T., Tanner, M.,
Hwang, S.-I., Polikoff, D., Nonet, G., Cochran, J., Myambo, K.,
Jay, K.E., Froula, J., Cloutier, T., Kuo, W.-L., Yaswen, P., Dairkee, S.,
Giovannola, J., Hutchinson, G.B., Isola, J., Kallioniemi, O.-P.,
Palazzolo, M., Martin, C., Ericsson, C., Pinkel, D., Albertson, D.,
Li, W.-B. and Gray, J.W.
TITLE Direct Submission
JOURNAL Submitted (07-JAN-1998) Life Sciences, Lawrence Berkeley National
```

```
FEATURES
source Laboratory, 1 Cyclotron Road, Berkeley, CA 94720, USA
Location/Qualifiers
1..2813
/organism="Homo sapiens"
/mol_type="mrna"
/db_xref="taxon:9606"
/chromosome="20"
/map="20q13.2-q13.3"
gene 1..2813
/feature="AIB1"
CDS 118..1872
/feature="AIB1"
/notes="amplified and overexpressed in breast cancer"
/codon_start=1
/product="AIB1"
/protein_id="AAC39896.1"
/db_xref="GI:3335398"
/translation="MGNQMSVQFQVDEQNEPEAETYNASALNGVPVVVSTHTVOH
LEEVDLGISVKTNVATSSPETTEISAVDANGKNGLEAKPEAPAKSRFFLMLSRP
VPGRTGQAADSLGSLVSKLDVSNKANKDPSESFTLPVAGAPGQDTKTPGHPAQ
DKVLSAARDPTLLPPTGAGGAPSPKDSFDFKFLDKGQKVPFGSQQAQAKRA
EHQDKVDEVPGLSQSDVDVAGKDI VDGKEKGQELGTADCSVPDPEGLTAKDDSO
AAALANNNSIMSFFKTLVSPNKAETKKDPEDTGAESPTTSADLKSKDKANFTSOETQ
GAGNSKGCNPSTGHTQSTVTPPAKGTKEKSGPTSLPLGKLFHKKSVKEDSVTGAE
ENVVCSPEVILIKSEVESALQTVLDNEGDAPEPTAKLKREESKPTSLMAFLRQM
SVKGDGITHSEINGKSSQTSDEKTTTPPEPTGAPQKQKGGSSKKKSAAE
MNFQKSNQKQAKPEAQCETQATVDTNSLQNGDKLQKPEKRRQSLGSGFFKGLGPKRML
DAQVQDTPVSGIPVGRPK"
ORIGIN
Query Match 23.2%; Score 167.6; DB 9; Length 2813;
Best Local Similarity 98.2%; Pred. No. 6.4e-36;
Matches 167; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Qy 419 AGGTGGTGTGTGAGTCACAGGAGATTTAAAGTCCAGGAAGTAGAATCAGCCTTAC 478
Db 1298 ATGTGGTGTGTGAGTCACAGGAGATTTAAAGTCCAGGAAGTAGAATCAGCCTTAC 1357
Qy 479 AAACAGTGGACCTCAACGAAGGAGATGCTGCACCTGAAACCCACGAAAGCGAAACTCAAAA 538
Db 1358 AAACAGTGGACCTCAACGAAGGAGATGCTGCACCTGAAACCCACGAAAGCGAAACTCAAAA 1417
Qy 539 GAGAGAAAGCAACCAAGAACCTCTCTGATGRCGTTTCTCAGACAAATG 588
Db 1418 GAGAGAAAGCAACCAAGAACCTCTCTGATGRCGTTTCTCAGACAAATG 1467
|||||
RESULT 10
BD085735 3066 bp DNA linear PAT 27-AUG-2002
LOCUS
DEFINITION Genes from the 20q13 amplicon and their uses.
ACCESSION BD085735
VERSION BD085735.1 GI:22631345
KEYWORDS JP 2001524802-A/11.
SOURCE synthetic construct
ORGANISM synthetic construct
REFERENCE 1 (bases 1 to 3066)
AUTHORS Gray, J.W., Collins, C.C., Hwang, S.-I., Godfrey, T., Kowbel, D. and
Rommens, J.
TITLE Genes from the 20q13 amplicon and their uses
JOURNAL Patent: JP 2001524802-A 11 04-DEC-2001;
COMMENT THE REGENTS OF THE UNIVERSITY OF CALIFORNIA
OS Artificial Sequence
PN JP 2001524802-A/11
PD 04-DEC-2001
PR 15-JUL-1997 JP 1998506264
PR 15-JUL-1996 US 08/680395, 16-OCT-1996 US 08/731499 PR
17-JAN-1997 US 08/785532
PI JOE W GRAY, COLIN CONRAD COLLINS, SOO IN HWANG, TONY GODFREY, PI
DAVID KOWBEL,
PI JOHANNA ROMMENS
PC C12N15/11.C12Q1/68,A61K48/00
```

```

CC Description of Artificial Sequence:lb1
FH Key Location/Qualifiers
FT source 1..3066
FT /organism='Artificial Sequence'
FEATURES
    source
        Location/Qualifiers
            1..3066
            /organism="synthetic construct"
            /mol_type="genomic DNA"
            /db_xref="taxon:34630"
ORIGIN
Query Match 23.2%; Score 167.6; DB 6; Length 3066;
Best Local Similarity 98.2%; Pred. No. 6.4e-36;
Matches 167; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 419 AGTGTGTGTGAGTCACAGTAGAGATATATAAGTCCAAAGGAGTAGAATCAGCCTTAC 478
DB 1433 ATGTGTGTGTGAGTCACAGTAGAGATATATAAGTCCAAAGGAGTAGAATCAGCCTTAC 1492
QY 479 AAACAGTGGACCTCAACGAAGGAGATGCTGCACCTGAACCCAGCAAGCAAACTCAAAA 538
DB 1493 AAACAGTGGACCTCAACGAAGGAGATGCTGCACCTGAACCCAGCAAGCAAACTCAAAA 1552
QY 539 GAGAAGAAAGCAAAACCAAGAACTCTCTCTGATGCGTTTCTCAGACAAATG 588
DB 1553 GAGAAGAAAGCAAAACCAAGAACTCTCTCTGATGCGTTTCTCAGACAAATG 1602
RESULT 11
AC126535/c
LOCUS
DEFINITION
Rattus norvegicus clone CH230-408J13, *** SEQUENCING IN PROGRESS
*** 3 unordered pieces.
AC126535
VERSION
HTG; HTGS PHASE1; HTGS DRAFT; HTGS_ENRICHED.
KEYWORDS
Rattus norvegicus (Norway rat)
SOURCE
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 168876)
Muzny,D.,Marie,,Metzker,M.Lee,,Abramson,S.,Adams,C.,Alder,J.,
Allen,C.,Allen,H.,Alsbrooks,S.,Amin,A.,Anguiano,D.,
Anyalebechi,V.,Aoyagi,A.,Ayodeji,M.,Baca,E.,Baden,H.,
Baldwin,D.,Bandaranka,D.,Barber,M.,Barnstead,M.,Benahmed,F.,
Biswal,K.,Blair,J.,Blankenburg,K.,Blyth,P.,Brown,M.,
Bryant,N.,Buhay,C.,Burch,P.,Burrell,K.,Calderson,E.,
Cardenas,V.,Carter,K.,Cavazos,I.,Ceasar,H.,Center,A.,
Chacko,J.,Chavez,D.,Chen,G.,Chen,R.,Chen,Y.,Chen,Z.,Chu,J.,
Cleaveland,C.,Cockrell,R.,Cox,C.,Coyle,M.,Cree,A.,D'Souza,L.,
Davila,M.L.,Davis,C.,Davy-Carroll,L.,De Anda,C.,Dederich,D.,
Delgado,O.,Denson,S.,Derano,C.,Ding,Y.,Dinh,H.,Divya,K.,
Draper,H.,Dugan-Rocha,S.,Dunn,A.,Durbini,K.,Duval,B.,Eaves,K.,
Egan,A.,Escotto,M.,Eugene,C.,Evans,C.A.,Falls,T.,Fan,G.,
Fernandez,S.,Finley,M.,Flagg,N.,Forbes,L.,Foster,M.,Foster,P.,
Fraser,C.M.,Gabisai,A.,Ganta,R.,Garcia,A.,Garner,T.,Garza,M.,
Guebrgeorgis,E.,Geer,K.,Gill,R.,Grady,M.,Guerra,W.,Guevara,W.,
Gunaratne,P.,Haaland,W.,Hamil,C.,Hamilton,C.,Hamilton,K.,
Harvey,Y.,Havlak,P.,Hawes,A.,Henderson,N.,Hernandez,J.,
Hernandez,R.,Hines,S.,Hladun,S.L.,Hodgson,A.,Hogues,M.,
Hollins,B.,Howells,S.,Hulyk,S.,Hume,J.,Idlebird,D.,Jackson,A.,
Jackson,L.,Jacob,L.,Jiang,H.,Johnson,B.,Johnson,R.,Jolivet,A.,
Karpach,S.,Kelly,S.,Kelly,S.,Khan,Z.,King,L.,Kovar,C.,
Kowis,C.,Kraft,C.L.,Lebow,H.,Levan,J.,Lewis,L.,Li,Z.,Liu,J.,
Liu,J.,Liu,W.,Liu,Y.,London,P.,Longacre,S.,Lopez,J.,
Lorensheewa,L.,Loulsegh,H.,Lozada,R.J.,Lu,X.,Ma,J.,
Maheshwari,M.,Mahindartne,M.,Mahmoud,M.,Malloy,K.,Mangum,A.,
Mangum,B.,Mapua,P.,Martin,K.,Martin,R.,Martinez,E.,
Mawhney,S.,McLeod,M.P.,McNeill,T.Z.,Meenen,E.,
Milosavljevic,A.,Miner,G.,Minja,E.,Montemayor,J.,Moore,S.,
Morgan,M.,Morris,K.,Morris,S.,Munidasa,M.,Murphy,M.,Nair,L.,
Nankervis,C.,Neal,D.,Newton,N.,Nguyen,N.,Norris,S.,

```

```

Nwackeleleh,O., Okwuonu,G., Olarnpunaagoon,A., Pal,S., Parks,K.,
Pasternak,S., Paul,H., Perez,A., Perez,L., Pfankoch,C.,
Plapper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
Puzio,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
Sanders,W., Saverty,G., Scherer,S., Scott,G., Shatman,S., Shen,H.,
Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Umanik,K.,
Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
Weinstock,G. and Gibbs,R.A.
Direct Submission
Unpublished
2 (bases 1 to 168876)
Worley,K.C.
Direct Submission
Submitted (06-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 168876)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (19-SEP-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Sep 19, 2002 this sequence version replaced gi:21703469.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). As a result, the
sequence may extend beyond the ends of the clone and there may be
contigs that consist entirely of whole genome shotgun sequence
reads. Both end sequences and whole genome shotgun sequence only
contigs will be indicated in the feature table.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GPPD
Center clone name: CH230-408J13
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 151329 bases at least Q40
Consensus quality: 153110 bases at least Q30
Consensus quality: 154290 bases at least Q20
Estimated insert size: 162792; sum-of-contigs estimation
Quality coverage: 5x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 12922: contig of 12922 bp in length
* 12923 13022: gap of unknown length
* 13023 155459: contig of 142437 bp in length
* 155460 155559: gap of unknown length
* 155560 168876: contig of 13317 bp in length.
* Location/Qualifiers
1..168876
FEATURES
source

```

```

/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-408J13"
11757..12922
/notes="wgs_contig"
13023..14162
/notes="wgs_contig"
15743..16838
/notes="wgs_contig"

ORIGIN
Query Match      19.1%; Score 137.8; DB 2; Length 168876;
Best Local Similarity 73.8%; Pred. NO. 1.3e-27;
Matches 172; Conservative 2; Mismatches 59; Indels 0; Gaps 0;

Qy 369 TTGGGGGAATGTTTGTGTTTTCATTTCTTATTACCTGGCAAAATAATCCAGGTGGTGTG 428
Db 143490 TGGGGTGTGTTTGTGTTTCTTATTACCTGGCAAAACACTCTAGGCGGTGTG 143431

Qy 429 TGAGTCCAGCTAGAGATTAAAGTCCAGGAAGTAGAATCAGCTTACAAAAGTGGGA 488
Db 143430 TGAGCCACCTAGAGACTGTAAGGCTTGAGGAAGTAGAATCCACCTTACAAACCGTGA 143371

Qy 489 CCTCAACGAGAGATGCTGCACCTGACCCACGACGCAAACTCAAAAGAGAAAG 548
Db 143370 TCTCACTGAGAAAGAAACCCAGACTGAGCCACAGATGTGAAGTCAAAGAGAAAGCAA 143311

Qy 549 CAACACCAAGAACCTCTCGATGTCGTTTCTCAGACAAATGTAAGCCCTTAC 601
Db 143310 ACCCGGAGGACCCCTCGATGTCGTTTCTCAGACAAATGTAAGCCACTTCC 143258

RESULT 12
AC094145/c. 233233 bp DNA linear HTG 09-MAY-2003
LOCUS Rattus norvegicus clone CH230-2L13, WORKING DRAFT SEQUENCE, 6
DEFINITION unordered pieces.
ACCESSION AC094145
VERSION AC094145.6 GI:30467887
KEYWORDS HTG; HTGS PHASE1; HTGS DRAFT; HTGS_FULLTOP.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 233233)
AUTHORS Muzny D.Marie., Metzker M.Lee., Abramson S., Adams C., Alder J.,
Allen C., Allen H., Alsbrooks S., Amin A., Anguiano D.,
Anyalebechi V., Aoyagi A., Ayodeji M., Baca E., Baden H.,
Baldwin D., Bandaranaike D., Barber M., Barnstead M., Benahmed F.,
Biswal K., Blair J., Blankenburg K., Blyth P., Brown M.,
Bryant N., Buhay C., Burch P., Burrell K., Calderon E.,
Cardenas V., Carter K., Cavazos I., Ceasar H., Center A.,
Chacko J., Chavez D., Chen G., Chen R., Chen Y., Chen Z., Chu J.,
Cleveland C., Cockrell R., Cox C., Coyle M., Cree A., D'Souza L.,
Davila M.L., Davis C., Davy-Carroll L., De Anda C., Dederich D.,
Delgado O., Denson S., Deramo C., Ding Y., Dinh H., Divya K.,
Draper H., Dugan-Rocha S., Dunn A., Durbin K., Duval B., Eaves K.,
Egan A., Escotto M., Eugene C., Evans C.A., Falls T., Fan G.,
Fernandez S., Finley M., Flagg N., Forbes L., Foster M., Foster P.,
Fraser C.M., Gabisi A., Ganta R., Garcia A., Garner T., Garza M.,
Gebregorgis E., Geer K., Gill R., Grady M., Guerra R., Guevara W.,
Gunaratne P., Haaland W., Hamil C., Hamilton C., Hamilton K.,
Harvey Y., Havlak P., Hawes A., Henderson N., Hernandez J.,
Hernandez R., Hines S., Hladun S.L., Hodgson A., Hognes M.,
Hollins B., Howells S., Hulyk S., Hume J., Idlebird D., Jackson A.,
Jackson L., Jacob L., Jiang H., Johnson B., Johnson R., Jolivet A.,
Karpathy S., Kelly S., Kelly S., Khan Z., King L., Kovar C.,
Kovis C., Kraft C.L., Lebow H., Levan J., Lewis L., Li Z., Liu J.,
Liu J., Liu W., Liu Y., London P., Longacre S., Lopez J.,
Lorensuhewa L., Loulisedge H., Lozano R.J., Lu X., Ma J.,
Maheshwari M., Mahindartne M., Mahmoud M., Malloy K., Mangum A.,

```

```

Mangum B., Mapua P., Martin K., Martin R., Martinez E.,
Mawhney S., McLeod M.P., McNeill T.Z., Meenen E.,
Milosavljevic A., Miner G., Minja E., Montemayor J., Moore S.,
Morgan M., Morris K., Morris S., Muidasa M., Murphy M., Nair L.,
Nankervis C., Neal D., Newton N., Nguyen N., Norris S.,
Nwaokemeleh O., Okwuonu G., Olarnpunsagoon A., Pal S., Parks K.,
Pasternak S., Paul H., Perez A., Perez L., Pfannkuch C.,
Popper F., Poindexter A., Popovic D., Primus E., Pu L.-L.,
Puzo M., Quiroz J., Rachlin E., Reeves K., Regier M.A., Reigh R.,
Reilly B., Reilly M., Ren Y., Reuter M., Richards S., Riggs F.,
Rives C., Rodkey T., Rojas A., Rose M., Rose R., Ruiz S.J.,
Sanders W., Savary G., Scherer S., Scott G., Shateman S., Shen H.,
Shetty J., Shvartsbeyn A., Sisson I., Sitter C.D., Smajda D.,
Sneed A., Sodergren E., Song X.-Z., Sorelle R., Sosa J.,
Steinle M., Strong R., Sutton A., Svatek A., Taber P., Taylor C.,
Taylor T., Thomas N., Thomas S., Tingey A., Trejos Z., Usmani K.,
Valas R., Vera V., Villalana D., Waldron L., Walker B., Wang J.,
Wang O., Wang S., Warren J., Warren R., Wei X., White F.,
Williams G., Willson R., Wlezyk R., Wooden H., Worley K.,
Wright D., Wright R., Wu J., Yakub S., Yen J., Yoon L., Yoon V.,
Yu F., Zhang J., Zhou J., Zhou X., Zhao S., Dunn D., von
Niederhausern A., Weiss R., Smith D.R., Holt R.A., Smith H.O.,
Weinstock G. and Gibbs R.A.
Direct Submission
Unpublished
2 (bases 1 to 233233)
Worley K.C.
Direct Submission
Submitted (17-SEP-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 233233)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (09-MAY-2003) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On May 9, 2003 this sequence version replaced gi:24942806.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GABR
Center clone name: CH230-2L13
----- Summary Statistics
Assembly program: Atlas;
Consensus quality: 207295 bases at least Q40
Consensus quality: 210461 bases at least Q30
Consensus quality: 212871 bases at least Q20
Estimated insert size: 217628; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.

```

* This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

* 1 89147: contig of 89147 bp in length
 * 89148 89247: gap of unknown length
 * 89248 99513: contig of 10266 bp in length
 * 99514 99613: gap of unknown length
 * 99614 113910: contig of 14297 bp in length
 * 113911 114010: gap of unknown length
 * 114011 230487: contig of 116477 bp in length
 * 230488 230587: gap of unknown length
 * 230588 232001: contig of 1414 bp in length
 * 232002 232101: gap of unknown length
 * 232102 232323: contig of 1132 bp in length.

FEATURES

source

1. .232323
 /organism="Rattus norvegicus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10116"
 /clone="CH230-2L13"

misc_feature

1. 1412
 /note="wgs_contig"

ORIGIN

Query Match 19.1%; Score 137.8; DB 2; Length 232323;
 Best Local Similarity 73.8%; Pred. No. 1.2e-27;
 Matches 172; Conservative 2; Mismatches 59; Indels 0; Gaps 0;
 QY 369 TTGGGGGAATGTTTCTTATTTCTTATTTACCTGGCAAAATATCCAGTGTGTG 428
 Db 77599 TGGGGTGTGTTTGTGTTTGTGTTTCTTATTTACCTGGCAAACTCTAGCGGTG 77540
 QY 429 TGAGTCACAGTAGATTATAAGTCCAGGAAGTAGATCAGCTTACAACTGCA 488
 Db 77539 TGAGCCACAGTAGAGACTGTAAAGCTTGAGGAAGTAGATCCACCTTACAAACCGTGA 77480
 QY 489 CCTCAACGAAGGAGATGCTGCACCTGAAACCCACGAAAGCAAACTCAAAAGAGAAGAAAG 548
 Db 77479 TCTCACTGAGAAAGAACCCAGACTGAGCCACAGATGTGAAGTCAAAAGAGAAACGCA 77420
 QY 549 CAACCAAGAACCTCTCTGATGRCGTTTCTCAGACAAATGGTAAGCCCTTAC 601
 Db 77419 ACCCGGAGGAGCCCTCTGTATGCGTTTCTCAGACAAATGGTAAGCCACTTC 77367

RESULT 13

AL928812/c

LOCUS Mouse DNA sequence from clone RP23-321B23 on chromosome 2, complete
 DEFINITION sequence.

ACCESSION AL928812

VERSION AL928812.11 GI:31335541

KEYWORDS HTG.

SOURCE Mus musculus (house mouse)

ORGANISM

Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 199019)

AUTHORS Heath, P.

TITLE Direct Submission

JOURNAL Submitted (30-MAY-2003) Wellcome Trust Sanger Institute, Hinxton,
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
 humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk

COMMENT

On Jun 2, 2003 this sequence version replaced gi:31096514.
 Sequence from the Mouse Genome Sequencing Consortium whole genome
 shotgun may have been used to confirm this sequence. Sequence data
 from the whole genome shotgun alone has only been used where it has
 a phred quality of at least 30.

----- Genome Center

Center: Wellcome Trust Sanger Institute

Center code: SC

Web site: http://www.sanger.ac.uk

Contact: humquery@sanger.ac.uk

 During sequence assembly data is compared from overlapping clones.
 Where differences are found these are annotated as variations
 together with a note of the overlapping clone name. Note that the
 variation annotation may not be found in the sequence submission
 corresponding to the overlapping clone, as we submit sequences with
 only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all
 regions were either double-stranded or sequenced with an alternate
 chemistry or covered by high quality data (i.e., phred quality >=
 30); an attempt was made to resolve all sequencing problems, such
 as compressions and repeats; all regions were covered by at least
 one plasmid subclone or more than one M13 subclone; and the
 assembly was confirmed by restriction digest, except on the rare
 occasion of the clone being a YAC.

The following abbreviations are used to associate primary accession
 numbers given in the feature table with their source databases:
 Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information
 on the WORMPEP database can be found at
 http://www.sanger.ac.uk/projects/C_elegans/wormpep RP23-321B23 is
 from the RPCI-23 Mouse BAC Library

constructed by the group of Pieter de Jong.

For further details see http://www.chori.org/bacpac/home.htm

VECTOR: pBACE3.6.

Location/Qualifiers

source

1. .199019
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10090"
 /chromosome="2"
 /clone="RP23-321B23"
 /clone_lib="RPCI-23"

ORIGIN

Query Match 17.6%; Score 127.4; DB 10; Length 199019;
 Best Local Similarity 74.7%; Pred. No. 1.1e-24;
 Matches 171; Conservative 2; Mismatches 53; Indels 3; Gaps 1;
 QY 371 GGGGGAATCTTTGTTTTCATTTCTTATACCTGGCAAAATATCCAGTGTGTG 430
 Db 165827 GGGTGTGTTTGTGTTTGTGTTTCTTATACCTGGCAAAATATCTAGCGGTGTG 165768
 QY 431 AGTCACAGTAGAGATTATAAGTCCAGGAAGTAGAATCAGCTTACAAACAGTGGACC 490
 Db 165767 AGTCACAGTAGAGACCGTAAGCTTGAGGAAGTAGAATCCAGCTTACAAACTGTGGATC 165708
 QY 491 TCACGAAGAGATGCTGCACCTGACCCACGACGAACTCAAAAGAGAGAAAGCA 550
 Db 165707 TCAGTGA---AGAGACCCAGCCTGAACCCACAGACGTAAGTCAAAAGAGAAAGCAAC 165651
 QY 551 AACCAAGAACCTCTCTGATGRCGTTTCTCAGACAAATGGTAAGCCCTT 599
 Db 165650 CCGGGAAGACCCCTCTGATGGCGTTTCTCAGACAAATGGTAAGCCACTT 165602

RESULT 14

AC084066/c

LOCUS

DEFINITION Mus musculus clone RP23-321D1, *** SEQUENCING IN PROGRESS ***, 29
 unsorted pieces.

ACCESSION AC084066

VERSION AC084066.1 GI:10799415

KEYWORDS HTG; HTGS, PHASE1.

SOURCE Mus musculus (house mouse)

ORGANISM

Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 235411)

AUTHORS DOE Joint Genome Institute.

TITLE Sequencing of Mouse

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 235411)

AUTHORS DOE Joint Genome Institute.

TITLE
JOURNAL
COMMENT

Direct Submission
Submitted (12-OCT-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: <http://www.jgi.doe.gov>

Project Information
Center Project Name: 2351294
Center clone name: RPCI-23_321D1

Summary Statistics

Consensus quality: 214207 bases at least Q40
Consensus quality: 223053 bases at least Q30
Consensus quality: 225208 bases at least Q20
Estimated insert size: 200000; pulse field gel estimation
Estimated insert size: 232611; sum-of-contigs estimation
Quality coverage: 11.56 in Q20 bases; pulse field gel estimation
Quality coverage: 9.94 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 29 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 1210: contig of 1210 bp in length
* 1211 1310: gap of unknown length
* 1311 2240: contig of 1330 bp in length
* 3241 3340: gap of unknown length
* 3341 4504: contig of 1164 bp in length
* 4505 4604: gap of unknown length
* 4605 6172: contig of 1568 bp in length
* 6173 6272: gap of unknown length
* 6273 7354: contig of 1082 bp in length
* 7355 7454: gap of unknown length
* 7455 8625: contig of 1171 bp in length
* 8626 8725: gap of unknown length
* 8726 10114: contig of 1389 bp in length
* 10115 10214: gap of unknown length
* 10215 12091: contig of 1877 bp in length
* 12092 12191: gap of unknown length
* 12192 14113: contig of 1922 bp in length
* 14114 14213: gap of unknown length
* 14214 15286: contig of 1073 bp in length
* 15287 15386: gap of unknown length
* 15387 17839: contig of 2453 bp in length
* 17840 17939: gap of unknown length
* 17940 20266: contig of 2327 bp in length
* 20267 20366: gap of unknown length
* 20367 22001: contig of 1635 bp in length
* 22002 22101: gap of unknown length
* 22102 24307: contig of 2206 bp in length
* 24308 24407: gap of unknown length
* 24408 26405: contig of 1998 bp in length
* 26406 26505: gap of unknown length
* 26506 29691: contig of 3186 bp in length
* 29692 29791: gap of unknown length
* 29792 34705: contig of 4914 bp in length
* 34706 34805: gap of unknown length
* 34806 39749: contig of 4944 bp in length
* 39750 39849: gap of unknown length
* 39850 45296: contig of 5447 bp in length
* 45297 45396: gap of unknown length
* 45397 51476: contig of 6080 bp in length
* 51477 51576: gap of unknown length
* 51577 59008: contig of 7432 bp in length
* 59009 59108: gap of unknown length
* 59109 66218: contig of 7110 bp in length
* 66219 66318: gap of unknown length
* 66319 76778: contig of 10460 bp in length

FEATURES
source
1. .235411
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/clone="RP23-321D1"
/clone_lib="RPCI mouse BAC library 23"

ORIGIN

Query Match 17.6%; Score 127.4; DB 2; Length 235411;
Best Local Similarity 74.7%; Pred. No. 1e-24;
Matches 171; Conservative 2; Mismatches 53; Indels 3; Gaps 1;
Qy 371 GGGGAATGTTTCTTTTCAATTTCTTATACCTGCGCAAAATATCCAGGTGTGTG 430
Db 225082 GGGTGTGTTTGTGTTTCTTATACCTGCGCAAAATATCTAGGCGGTGTG 225023
Qy 431 AGTCACAGTAGAGATTATAAGTCCAAAGGAGTAGAATCAGCTTACAAACAGTGGACC 490
Db 225022 AGTCACAGTAGAGACCGTAAGGCTTGAGGAGTAGAATCAGCTTACAACTGGATC 224963
Qy 491 TCAACGAAGGAGATGCTGCACCTGAACCCACGAAAGCGAAACTCAAAAGAGAAGCA 550
Db 224962 TCAGTGA---AGACACCCAGCCTCAACCCACAGACGTAAGAGTCAAGAGCAAAAC 224906
Qy 551 AACCAAGAACTCTCTGATGCGTTTCTCAGACAAATGGTAAGCCCTT 599
Db 224905 CCCGAAGACCCCTCTGATGCGTTTCTCAGACAAATGGTAAGCCACTT 224857

RESULT 15

AY219233 2916 bp mRNA linear ROD 20-OCT-2003
LOCUS Mus musculus novel amplified in breast cancer-1 mRNA, complete cds.
DEFINITION
ACCESSION AY219233
VERSION AY219233.1 GI:29648617
KEYWORDS
SOURCE Mus musculus (house mouse)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 2916)
AUTHORS Beardsley, D.I., Kowbel, D., Lataxes, T.A., Mannino, J.M., Xin, H.,
Kim, W.J., Collins, C. and Brown, K.D.
TITLE Characterization of the novel amplified in breast cancer-1 (NABCL)
gene product
JOURNAL Exp. Cell Res. 290 (2), 402-413 (2003)
MEDLINE 22930202
PUBMED 14567997
REFERENCE 2 (bases 1 to 2916)
AUTHORS Brown, K.
TITLE Direct Submission
JOURNAL Submitted (14-JAN-2003) Biochem. and Molec. Biology, LSU Health
Sciences Center, 1901 Perdido Street, New Orleans, LA 70112, USA
FEATURES
source
1. .2916
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
89..1990
/note="NABCL; AIBCL; BCAS1"

CDS

```
/codon_start=1
/product="novel amplified in breast cancer-1"
/protein_id="AA088011.1"
/db_xref="GI:29648618"
/translation="MGNQMSVPLRPGDQHDGADTKVTSNECVQNGNPVVLSTRV
IQHYEVDIGISSKDNVATSPFKMEQAQVADSGKNGLKAEAKTKAPAARSHFFLTL
SRPVPGRPDQGTDSAAQSERFDVPSAADPNKDPSEHGALFVAAPQAQDKTPCCP
EAKQOTLPATGLAPSPPEQAEPQAQDKDFGLNRFKLDKGRESAPVNSQPKAAG
SDPQATAPAVPGNPHGVSAQEDIVDSEQRQDVDTLSYVPGDPEVPGTTKEDPQ
VVDTTNSSIMSFFKTLVSPNKTETKDPEDTKATKADSVCDGHAAGQKMSQTQAKS
KKRLDSPRLGUSFRKLFRRKDTENSPTTSANLKDKNFTPOETRGKTKATKSCPP
PPSPFTSGRSGSGKEKAGPTLLPLGKLFWKSVKEDTLSTGAENAVCESPEVTVRL
EEVSSLQTVDLSEETQPEPTDVKVEESKPRKPLMAFLQMSVRSSEGI PRSEEN
VKDSSCOTNSVEKTPSPPEPPAGTAQKNKETSSKDKKSVDKKSATENSQKNGKQ
EVREPAPCVQPTVEANAMQTGDKTPKSEKRQSLGGLKGLGPKRMSDAQVQTDPV
SIGPVGKSK"
```

ORIGIN

Query Match		12.0%; Score 86.6; DB 10; Length 2916;
Best Local Similarity		60.7%; Pred. No. 3.9e-13;
Matches	156; Conservative	2; Mismatches 96; Indels 3; Gaps 1;
QY	332	AGGGGGGAGCAAGGACCCACTTTGTAAACATGTTTGGGGGAATGTTTGTTCATCA 391
Db	1314	AGGCAGGGCCCACTTACTACCGCTGGGAAAGTTGTTTGGAGAAAGTCAGTTAAGGAGG 1373
QY	392	TTTTCTTATTACCTGCAAAATAATCCAGTGTGTGAGTCACCAGTAGAGATTATAA 451
Db	1374	ATACACTTTCCACAGGTGCAGAGGAGAACCGGTGTGTGAGTCACCAGTAGAGACCGTAA 1433
QY	452	AGTCCAAGGAAGTAGAATCAGCCTTACAAACAGTGGACCTCAACGAAAGGAGATGCTGCAC 511
Db	1434	GGCTTCAGGAAGTAGAATCCAGCTTACAACTGTGGATCTCAGTGA--AGAGACCCAGC 1490
QY	512	CTGAACCCACWGAAGCGAAACTCAAAAGAGAGAAAGCAAAACCAAGAACTCTCTGATGR 571
Db	1491	CTGAACCCACAGACGTAAAAGTCAAGAGAGAAAGCAAAACCCCGAAGACCCCTCTGATGG 1550
QY	572	CGTTTCTCAGACAAATG 588
Db	1551	CGTTTCTCAGACAAATG 1567

Search completed: May 5, 2005, 03:58:10
Job time : 3543.14 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 4, 2005, 20:02:41 ; Search time 500.471 Seconds
(without alignment)
8551.893 Million cell updates/sec

Title: US-08-731-499-2

Perfect score: 723

Sequence: 1 TGGAGCTGTCATGCTTACC.....TGAATGGATATGATTCG 723

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseqn_16Dec04.*

1: Geneseqn1980s.*

2: Geneseqn1990s.*

3: Geneseqn2000s.*

4: Geneseqn2001as.*

5: Geneseqn2001bs.*

6: Geneseqn2002as.*

7: Geneseqn2002bs.*

8: Geneseqn2003as.*

9: Geneseqn2003bs.*

10: Geneseqn2003cs.*

11: Geneseqn2003ds.*

12: Geneseqn2004as.*

13: Geneseqn2004bs.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	722.2	99.9	723	2	AAV04697 Homo sapi
2	167.6	23.2	349	2	AAV68411 EST clone
3	167.6	23.2	2813	5	AA581115 DNA encod
4	167.6	23.2	2813	6	AA38092 Human BCA
5	167.6	23.2	2813	12	ADJ75127 Marker ge
6	167.6	23.2	2813	12	ADQ18993 Human sof
7	167.6	23.2	2813	13	ADR25076 Breast ca
8	167.6	23.2	3066	2	AAV09025 Homo sapi
9	167.6	23.2	3418	12	ADQ23347 Human sof
10	166	23.0	2020	8	ACC90605 Human CGD
11	88.6	12.3	383	8	ABX44866 Bovine ES
12	70.2	9.7	540	13	ADQ53268 Novel can
13	41.2	5.7	957	6	ABZ13040 Arabidops
14	41.2	5.7	1347	3	AAC50675 Arabidops
15	41.2	5.7	1352	3	AAC33342 Arabidops
16	41.2	5.7	1352	3	AAC50674 Arabidops
17	41.2	5.7	1356	3	AAC36193 Arabidops
18	41.2	5.7	1470	3	AAC50673 Arabidops
19	41.2	5.7	1474	3	AAC39251 Arabidops
20	39.4	5.4	103464	13	ABD33278 Murine ca

c

c	21	37.8	5.2	2000	8	ADA71938	Ada71938 Rice gene
c	22	37.2	5.1	17534	6	ABK40026	ABK40026 Human che
	23	37.2	5.1	83836	11	ACN45080	ACN45080 Mouse gen
	24	36.6	5.1	3352	5	ABV27806	ABV27806 Human pro
	25	36.6	5.1	3353	5	ABV25280	ABV25280 Human pro
	26	36.4	5.0	428	4	AAI88208	AAI88208 Human pol
	27	36.2	5.0	198522	11	ACN44010	ACN44010 Human gen
c	28	36	5.0	598	12	ACH67256	ACH67256 Human gen
	29	36	5.0	885	10	ADF68708	ADF68708 Human met
	30	36	5.0	885	10	ADF89281	ADF89281 P gene se
	31	36	5.0	885	12	ADJ97169	ADJ97169 P gene DN
	32	36	5.0	885	12	ADL07820	ADL07820 P gene fr
	33	36	5.0	885	12	ADM67640	ADM67640 Human met
	34	36	5.0	7734	4	ABL29804	ABL29804 Drosophil
	35	36	5.0	13215	10	ADF68721	ADF68721 Human met
	36	36	5.0	13215	10	ADF89222	ADF89222 Human met
	37	36	5.0	13215	12	ADJ96810	ADJ96810 Human met
	38	36	5.0	13215	12	ADL07833	ADL07833 HMPV cDNA
	39	36	5.0	13215	12	ADM67281	ADM67281 Human met
c	40	35.8	5.0	110000	12	ADN06353	ADN06353 Human FLA
c	41	35.8	5.0	110000	13	ADS94372	ADS94372 Human 5-1
	42	35.4	4.9	13021	4	AAI04666	AAI04666 Human rep
	43	35.4	4.9	13021	4	ABL97573	ABL97573 Human tes
c	44	35.4	4.9	26997	4	AA546747	AA546747 Tumour su
	45	35.2	4.9	552	5	ADL43717	ADL43717 Human ova

ALIGNMENTS

RESULT 1

AAV04697
ID AAV04697 standard; cDNA; 723 BP.

XX AAV04697;

XX 21-JUL-1998 (first entry)

XX Homo sapiens 20q13 amplicon 1b11 transcript.

XX 20q13 amplicon; chromosome 20; tumour; detection;
XX chromosomal abnormalities; probe; gene therapy; antisense inhibition;
XX treatment; age-related macular degeneration; retinitis pigmentation;
XX Leber's congenital amaurosis; ds.

XX Homo sapiens.

XX WO9802539-A1.

XX 22-JAN-1998.

XX 15-JUL-1997; 97WO-US012343.

XX 15-JUL-1996; 96US-00580395.

XX 16-OCT-1996; 96US-00731499.

XX 17-JAN-1997; 97US-00785532.

XX (REGC) UNIV CALIFORNIA.

XX Gray JW, Collins CC, Hwang S, Godfrey T, Kowbel D, Rommens J;

XX WPI; 1998-110587/10.

XX New sequences from the 20q13 amplicon - used for detecting chromosomal

XX abnormalities, particularly tumours, and for developing products for

XX treating diseases.

XX Claim 1; Page 61; 91pp; English.

XX The sequence is that of a cDNA sequence 1b11, which was isolated from the

XX 20q13 amplicon. Its expression shows high correlation with the copy

XX number of the amplicon. It can be used as a probe for the detection of

XX chromosomal abnormalities at 20q13. It and other sequences isolated from

CC the 20q13 amplicon are consistently amplified in primary tumours. These
 CC sequences are useful as probes or as probe targets for monitoring the
 CC relative copy number of corresponding sequences from a biological sample
 CC such as tumour cells. The sequences can also be used in therapeutic
 CC applications for modulating the expression of the endogenous gene or the
 CC activity of the gene product. Examples of therapeutic approaches include
 CC antisense inhibition of gene expression, gene therapy, and monoclonal
 CC antibodies that specifically bind the gene products. The products can
 CC also be used in the treatment of other diseases, e.g. age-related macular
 CC degeneration, Leber's congenital amaurosis and retinitis pigmentation
 XX
 SQ Sequence 723 BP; 210 A; 138 C; 178 G; 195 T; 0 U; 2 Other;

Query Match 99.9%; Score 722.2; DB 2; Length 723;
 Best Local Similarity 100.0%; Pred. No. 1.2e-222;
 Matches 723; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGAAGCTGTCTATGGTTACCGTCTTAACGTTGGACTCTTAAGAAATGATTATTCCTGG 60
 Db 1 TGAAGCTGTCTATGGTTACCGTCTTAACGTTGGACTCTTAAGAAATGATTATTCCTGG 60
 QY 61 TTCTAGACAGGCCCAATGTAATTCACCTACGTCGAGATTAAAGAGTGGGCTTACTAG 120
 Db 61 TTCTAGACAGGCCCAATGTAATTCACCTACGTCGAGATTAAAGAGTGGGCTTACTAG 120
 QY 121 ATTGATTGGGTATTGAGCATGCTCTGAATGACAGTCCCCAAAGAGCCTCTTATCCGT 180
 Db 121 ATTGATTGGGTATTGAGCATGCTCTGAATGACAGTCCCCAAAGAGCCTCTTATCCGT 180
 QY 181 TCTTCCCTTTGGGGAAGGGCTTTGGCACTTCCATGCAATGTCAGTGGAGTTGGAA 240
 Db 181 TCTTCCCTTTGGGGAAGGGCTTTGGCACTTCCATGCAATGTCAGTGGAGTTGGAA 240
 QY 241 ATTGGTGGCTGTACAACTAGCATTAATCTCTCAAGATGTCCTGTGTAGAAATGGTC 300
 Db 241 ATTGGTGGCTGTACAACTAGCATTAATCTCTCAAGATGTCCTGTGTAGAAATGGTC 300
 QY 301 ATAGATTCAAACTGTAGCTACTATGTGACAGGGGGGAGCAAGACCCACTTTGTAA 360
 Db 301 ATAGATTCAAACTGTAGCTACTATGTGACAGGGGGGAGCAAGACCCACTTTGTAA 360
 QY 361 AACATGTTTGGGGAATGTTTGTCTTATTTTCAATTTCTTATACCTGGCAAAATAATCCAG 420
 Db 361 AACATGTTTGGGGAATGTTTGTCTTATTTTCAATTTCTTATACCTGGCAAAATAATCCAG 420
 QY 421 GTGGTGTGAGTCACAGTAGAGATTATAAGTCCAAAGGAAGTAGAATCAGCCTTACAA 480
 Db 421 GTGGTGTGAGTCACAGTAGAGATTATAAGTCCAAAGGAAGTAGAATCAGCCTTACAA 480
 QY 481 ACAGTGGACCTCAACGAAGAGATGTCGACCTGAAACCCACGAAAGCAAACTCAAAAGA 540
 Db 481 ACAGTGGACCTCAACGAAGAGATGTCGACCTGAAACCCACGAAAGCAAACTCAAAAGA 540
 QY 541 GAAGAAGCAAAACCAAGAACTCTCTGATGRCGTTTCTCAGACAAATGGTAAGCCCTTA 600
 Db 541 GAAGAAGCAAAACCAAGAACTCTCTGATGRCGTTTCTCAGACAAATGGTAAGCCCTTA 600
 QY 601 CTTCCAGTATAGAAACCTTAAGATACCTAGAGGGGCTTTTGGGAAACAATGGGCTCATGCC 660
 Db 601 CTTCCAGTATAGAAACCTTAAGATACCTAGAGGGGCTTTTGGGAAACAATGGGCTCATGCC 660
 QY 661 ACAGGTAGTAGGAGACATAATTCTAGCTGGTGTGTAATGGAATCTGAATGGAATATGGATT 720
 Db 661 ACAGGTAGTAGGAGACATAATTCTAGCTGGTGTGTAATGGAATCTGAATGGAATATGGATT 720
 QY 721 GCG 723
 Db 721 GCG 723

RESULT 2
 AAV88411
 ID AAV88411 standard; cDNA; 349 BP.

XX AAV88411;
 AC
 XX
 DT 12-FEB-1999 (first entry)
 XX
 DE EST clone EK480.
 XX
 KW Expressed sequence tag; secreted protein; haematopoiesis regulator;
 KW tissue growth; activin; inhibin; tumour invasion suppressor; EST; human;
 KW chemotaxis; chemokinesis; haemostasis; gene therapy; thrombolysis;
 KW receptor; ligand; anti-inflammatory; tumour inhibitor; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO9845437-A2.
 XX
 PD 15-OCT-1998.
 XX
 PF 10-APR-1998; 98WO-US006956.
 XX
 PR 10-APR-1997; 97US-00837312.
 XX
 PA (GEMY) GENETICS INST INC.
 XX
 PI Jacobs K, McCoy JM, Lavallie ER, Racie LA, Merberg D, Treacy M;
 PI Spaulding V, Agostino MJ;
 XX
 DR WPI; 1999-070078/06.
 XX
 PT New polynucleotides encoding human secreted proteins - derived from e.g.
 PT human blood, kidney, foetal lung, placenta, testes, brain, ovary,
 PT pituitary, retina and colon cDNA libraries.
 XX
 PS Claim 1; Page 388; 641pp; English.
 XX
 CC The present sequence represents an expressed sequence tag (EST), and is a
 CC polynucleotide of the invention. The polynucleotides of the invention are
 CC all secreted EST sequences isolated from a variety of human tissue
 CC sources. The EST sequences and proteins encoded by them are predicted to
 CC have useful biological activities which would make them suitable for
 CC treating, preventing or ameliorating medical conditions in humans and
 CC animals, although no supporting data is given. Suggested activities
 CC include nutritional activity, immune stimulating or suppressing activity,
 CC haematopoiesis regulating activity, tissue growth activity,
 CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, receptor/ligand activity, anti-inflammatory
 CC activity, cadherin/tumour invasion suppressor activity, tumour inhibition
 CC activity. The EST sequences are also stated to be useful for gene therapy
 XX
 SQ Sequence 349 BP; 128 A; 81 C; 90 G; 50 T; 0 U; 0 Other;

Query Match 23.2%; Score 167.6; DB 2; Length 349;
 Best Local Similarity 98.2%; Pred. No. 2.5e-43;
 Matches 167; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 QY 419 AGTGTGTGTGTCAGTCACCAAGGAGATTAAGTCCAAAGTAGAATCAGCCTTAC 478
 Db 69 ATGTGTGTGTGTCAGTCACCAAGGAGATTAAGTCCAAAGTAGAATCAGCCTTAC 128
 QY 479 AAACAGTGACCTCAACGAAGGAGATGTCGACCTGAAACCCACGAAAGCAAACTCAAAA 538
 Db 129 AAACAGTGACCTCAACGAAGGAGATGTCGACCTGAAACCCACGAAAGCAAACTCAAAA 188
 QY 539 GAGAAGAAAGCAACCAAGAACTCTCTGATGRCGTTTCTCAGACAAATG 588
 Db 189 GAGAAGAAAGCAACCAAGAACTCTCTGATGRCGTTTCTCAGACAAATG 238
 RESULT 3
 AAS81115
 ID AAS81115 standard; cDNA; 2813 BP.
 XX
 AC AAS81115;

```
XX 13-FEB-2002 (first entry)
XX DNA encoding novel human diagnostic protein #16919.
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX Homo sapiens.
XX WO200175067-A2.
XX 11-OCT-2001.
XX 30-MAR-2001; 2001WO-US008631.
XX 31-MAR-2000; 2000US-00540217.
XX 23-AUG-2000; 2000US-00649167.
XX (HYSE-) HYSEQ INC.
XX Drmanac RT, Liu C, Tang YT;
XX WPI; 2001-639362/73.
XX P-PSDB; ASG16928.
XX New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits and to assess
XX biodiversity.
XX Claim 1; SEQ ID NO 16919; 103pp; English.
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
XX sequences. (I) is useful as hybridisation probes, polymerase chain
XX reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
XX and in recombinant production of (II). The polynucleotides are also used
XX in diagnostics as expressed sequence tags for identifying expressed
XX genes. (I) is useful in gene therapy techniques to restore normal
XX activity of (II) or to treat disease states involving (II). (II) is
XX useful for generating antibodies against it, detecting or quantitating a
XX polypeptide in tissue, as molecular weight markers and as a food
XX supplement. (II) and its binding partners are useful in medical imaging
XX of sites expressing (II). (I) and (II) are useful for treating disorders
XX involving aberrant protein expression or biological activity. The
XX polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
XX coding sequences of the invention. Note: The sequence data for this
XX patent did not appear in the printed specification, but was obtained in
XX electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX Sequence 2813 BP; 868 A; 671 C; 758 G; 516 T; 0 U; 0 Other;
Query Match 23.2%; Score 167.6; DB 5; Length 2813;
Best Local Similarity 98.2%; Pred. No. 7.6e-43;
Matches 167; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Qy 419 AGGTGGTGTGTGAGTCACAGTAGAGATTATAAAGTCCAGGAAGTAGAATCAGCCTTAC 478
Db 1298 ATGTGGTGTGTGAGTCACAGTAGAGATTATAAAGTCCAGGAAGTAGAATCAGCCTTAC 1357
Qy 479 AAACAGTGGACCTCAACGAAGGAGATGCTGCACCTGAACCCACGACGGAACCTCAAAA 538
Db 1358 AAACAGTGGACCTCAACGAAGGAGATGCTGCACCTGAACCCACGACGGAACCTCAAAA 1417
Qy 539 GAGAAGAAAGCAAAACCAAGAACCTCTCTGATGRCGTTTCTCAGACAAATG 588
Db 1418 GAGAAGAAAGCAAAACCAAGAACCTCTCTGATGRCGTTTCTCAGACAAATG 1467
```

```
RESULT 4
AAD38092
ID AAD38092 standard; DNA; 2813 BP.
XX
AC AAD38092;
XX
DT 10-SEP-2002 (first entry)
XX
DE Human BCAS1 (breast cancer amplified sequence 1) DNA.
XX
KW Human; BCAS1; breast cancer amplified sequence 1; AIBCL; inflammation;
KW amplified in breast cancer 1; NABCL; novel amplified in breast cancer 1;
KW hyperproliferative disorder; breast; prostate; cancer; prophylaxis;
KW infection; antisense therapy; cytostatic; antinflammatory; tumour; gene;
KW ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
CDS 118..1872
FT /*tag= a
FT /product= "Human BCAS1 protein"
XX
XX WO200231136-A1.
XX
PD 18-APR-2002.
XX
PF 09-OCT-2001; 2001WO-US031484.
XX
PR 11-OCT-2000; 2000US-00689255.
XX
PA (ISIS-) ISIS PHARM INC.
XX
PI Cowsett LM, Freier SM;
XX
DR WPI; 2002-444179/47.
XX
DR P-PSDB; AAE23791.
XX
XX New antisense compounds targeted to a nucleic acid molecule encoding
XX BCAS1, useful for treating diseases or conditions associated with BCAS1,
XX such as hyperproliferative disease, particularly breast or prostate
XX cancer.
XX Example 10; Page 92-95; 104pp; English.
XX The invention relates to antisense compounds, compositions and methods
XX for modulating the expression of BCAS1 (breast cancer amplified sequence
XX 1, also known as AIBCL for amplified in breast cancer 1 and NABCL for
XX novel amplified in breast cancer 1). The antisense compounds of the
XX invention are useful for treating an animal having a disease or condition
XX associated with BCAS1, such as hyperproliferative disorders including
XX breast or prostate cancer. These compounds are also used as research
XX reagents and diagnostics; to distinguish between functions of various
XX members of a biological pathway; in the treatment of a disease or
XX disorder, which can be treated by modulating the expression of BCAS1; as
XX prophylaxis, e.g. to prevent or delay infection, inflammation or tumour
XX formation, and as probes or primers. These antisense compounds are used
XX in antisense therapy. The present sequence is human BCAS1 DNA
XX
SQ Sequence 2813 BP; 868 A; 671 C; 758 G; 516 T; 0 U; 0 Other;
Query Match 23.2%; Score 167.6; DB 6; Length 2813;
Best Local Similarity 98.2%; Pred. No. 7.6e-43;
Matches 167; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Qy 419 AGGTGGTGTGTGAGTCACAGTAGAGATTATAAAGTCCAGGAAGTAGAATCAGCCTTAC 478
Db 1298 ATGTGGTGTGTGAGTCACAGTAGAGATTATAAAGTCCAGGAAGTAGAATCAGCCTTAC 1357
Qy 479 AAACAGTGGACCTCAACGAAGGAGATGCTGCACCTGAACCCACGACGGAACCTCAAAA 538
Db 1358 AAACAGTGGACCTCAACGAAGGAGATGCTGCACCTGAACCCACGACGGAACCTCAAAA 1417
```

QY 539 GAGAGAAAGCAAAACCAAGAACCTCTCTGATGRCGTTTCTCAGACAAATG 588
 DB 1418 GAGAGAAAGCAAAACCAAGAACCTCTCTGATGRCGTTTCTCAGACAAATG 1467

RESULT 5
 ADJ75127
 ID ADJ75127 standard; DNA; 2813 BP.
 XX AC ADJ75127;
 XX DT 20-MAY-2004 (first entry)
 XX DE Marker gene SEQ ID NO:379.
 XX KW bronchial asthma; chronic obstructive pulmonary disease;
 KW respiratory epithelial cell; interleukin-13; respiratory; antiasthmatic;
 KW gene therapy; marker gene; gene; ds.
 XX OS Homo sapiens.
 XX PN EP1394274-A2.
 XX PD 03-MAR-2004.
 XX PF 04-AUG-2003; 2003BP-00254857.
 XX PR 06-AUG-2002; 2002JP-00229312.
 XX PR 20-MAR-2003; 2003JP-00077212.
 XX PA (GENO-) GENOX RES INC.
 XX PI Ohtani N, Sugita Y, Yamaya M, Kubo H, Nagai H, Izuhara K;
 XX WPI; 2004-193155/19.
 XX Testing for bronchial asthma or chronic obstructive pulmonary disease by
 PT comparing the expression level of a marker gene in a biological sample
 PT from a subject with the expression level of the gene in a sample from a
 PT healthy subject.
 XX Claim 1; SEQ ID NO 379; 241pp; English.
 XX The present invention describes a method of testing for bronchial asthma
 CC or chronic obstructive pulmonary disease. The method comprises
 CC determining the expression level of a marker gene in a biological sample
 CC from a subject, comparing the expression level determined with the
 CC expression level of the marker gene in a biological sample from a healthy
 CC subject, and judging whether the subject has bronchial asthma or chronic
 CC obstructive pulmonary disease. The marker gene comprises: (a) a group of
 CC genes (S1) whose expression levels increase when respiratory epithelial
 CC cells are stimulated with interleukin-13; or (b) a group of genes (S2)
 CC whose expression levels decrease when respiratory epithelial cells are
 CC stimulated with interleukin-13. Also described: (1) a reagent (I) for
 CC testing for bronchial asthma or chronic obstructive pulmonary disease;
 CC (2) a kit for screening for a candidate compound for a therapeutic agent
 CC to treat bronchial asthma or chronic obstructive pulmonary disease; (3)
 CC an animal model for bronchial asthma or chronic obstructive pulmonary
 CC disease; (4) an inducer that induces bronchial asthma in a mouse; (5) a
 CC method for producing an animal model for bronchial asthma or chronic
 CC obstructive pulmonary disease; (6) a therapeutic agent for bronchial
 CC asthma or chronic obstructive pulmonary disease, comprising the compound,
 CC a marker gene or an antisense nucleic acid corresponding to a portion of
 CC the marker gene, a ribozyme, a polynucleotide that suppresses the
 CC expression of the gene through an RNAi effect or an antibody recognising
 CC a protein encoded by a marker gene; and (7) a DNA chip for testing for
 CC bronchial asthma or a chronic obstructive pulmonary disease, on which a
 CC probe has been immobilised to assay a marker gene. (I) has respiratory
 CC and antiasthmatic activities, and can be used in gene therapy. The method
 CC is useful for testing for or screening for a therapeutic agent for
 CC bronchial asthma or chronic obstructive pulmonary disease. The present
 CC sequence is used in the exemplification of the present invention.

XX SQ Sequence 2813 BP; 868 A; 671 C; 758 G; 516 T; 0 U; 0 Other;
 Query Match 23.2%; Score 167.6; DB 12; Length 2813;
 Best Local Similarity 98.2%; Pred. No. 7.6e-43;
 Matches 167; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 QY 419 AGTGGTGTGTGAGTCACAGTAGAGATTATAAATCTCAAGAGTAGAATCAGCCTTAC 478
 DB 1298 ATGTGGTGTGTGAGTCACAGTAGAGATTATAAATCTCAAGAGTAGAATCAGCCTTAC 1357
 QY 479 AAACAGTGGACCTCAACGAAGGAGATGCTGCACCTGAACCCACGAAAGCAAACTCAAAA 538
 DB 1358 AAACAGTGGACCTCAACGAAGGAGATGCTGCACCTGAACCCACGAAAGCAAACTCAAAA 1417
 QY 539 GAGAGAAAGCAAAACCAAGAACCTCTCTGATGRCGTTTCTCAGACAAATG 588
 DB 1418 GAGAGAAAGCAAAACCAAGAACCTCTCTGATGRCGTTTCTCAGACAAATG 1467

RESULT 6
 ADQ18993
 ID ADQ18993 standard; DNA; 2813 BP.
 XX AC ADQ18993;
 XX DT 26-AUG-2004 (first entry)
 XX DE Human soft tissue sarcoma-upregulated DNA - SEQ ID 1812.
 XX KW soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human;
 KW ds.
 XX OS Homo sapiens.
 XX PN WO2004048938-A2.
 XX PD 10-JUN-2004.
 XX PF 26-NOV-2003; 2003WO-US038193.
 XX PR 26-NOV-2002; 2002US-0429739P.
 XX PA (PROT-) PROTEIN DESIGN LABS INC.
 XX PI Aziz N, Ginsburg WM, Zlotnik A;
 XX WPI; 2004-441208/41.
 XX Early detection of soft tissue sarcoma comprises determining expression
 PT of a gene in a first soft tissue sample and a normal soft tissue sample
 PT and comparing the gene expression, also useful in treating soft tissue
 PT sarcoma.
 XX Example 2; SEQ ID NO 1812; 210pp; English.
 XX The invention relates to a novel method for detecting soft tissue sarcoma
 CC which comprises obtaining a first soft tissue sample from an individual
 CC and a normal soft tissue sample from the same or different individual,
 CC determining the expression of a gene in both samples and comparing the
 CC expression of the gene in both soft tissue samples, where a higher level
 CC of protein expression in the first soft tissue sample indicates the
 CC presence of soft tissue sarcoma. The method of the invention has
 CC cytostatic applications and may be useful for detecting soft tissue
 CC sarcoma, possibly via gene therapy or vaccine production. The nucleic
 CC acid sequences may be useful in diagnostic and screening applications.
 CC The current sequence is that of a human soft tissue sarcoma-upregulated
 CC DNA of the invention. The current sequence is not shown within the
 CC specification per se but was submitted in CD format by the inventor.
 XX SQ Sequence 2813 BP; 868 A; 671 C; 758 G; 516 T; 0 U; 0 Other;
 Query Match 23.2%; Score 167.6; DB 12; Length 2813;

```
Best Local Similarity 98.2%; Pred. No. 7.6e-43;
Matches 167; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 419 AGGTGGTGTGAGTCACCAAGAGAGATGCTGACCTGAAACCCACGAAAGCGAACTCAAAA 478
DB 1298 ATGTGGTGTGAGTCACCAAGAGAGATGCTGACCTGAAACCCACGAAAGCGAACTCAAAA 1357

QY 479 AAACAGTGGACCTCAACGAAGAGAGATGCTGACCTGAAACCCACGAAAGCGAACTCAAAA 538
DB 1358 AAACAGTGGACCTCAACGAAGAGAGATGCTGACCTGAAACCCACGAAAGCGAACTCAAAA 1417

QY 539 GAGAAGAAAGCAAAACCAAGAGAGATGCTGATGCGGCTTCTCAGACAAATG 588
DB 1418 GAGAAGAAAGCAAAACCAAGAGAGATGCTGATGCGGCTTCTCAGACAAATG 1467

RESULT 7
ID ADR25076 standard; DNA; 2813 BP.
XX
AC ADR25076;
XX
DT 21-OCT-2004 (first entry)
XX
DE Breast cancer prognosis marker #937.
XX
KW ds; breast cancer; prognosis; gene expression; diagnosis.
XX
OS Homo sapiens.
XX
FN WO2004065545-A2.
XX
PD 05-AUG-2004.
XX
PF 15-JAN-2004; 2004WO-US001100.
XX
PR 15-JAN-2003; 2003US-00342887.
XX
PA (ROSE-) ROSETTA INPHARMATICS LLC.
PA (NECA-) NETHERLANDS CANCER INST.
XX
PI Van't Veer LJ, He Y;
XX
DR WPI; 2004-593473/57.
XX
PS Classifying a breast cancer patient according to prognosis comprises
PT determining the similarity between the level of expression of each of
PT five genes in a cell sample taken from patient, to control levels.
XX
PS Disclosure; SEQ ID NO 937; 226pp; English.
XX
CC The invention relates to a method of classifying a breast cancer patient
CC according to prognosis by determining the similarity between the level of
CC expression of each of five genes for which markers are listed in the
CC specification, in a cell sample taken from the breast cancer patient, to
CC control levels of expression for each respective five genes to obtain a
CC patient similarity value. The methods are useful for classifying a breast
CC cancer patient according to prognosis. Kits and computer program products
CC are useful for data analysis using the diagnostic, prognostic and
CC statistical methods of the invention. This sequence corresponds to a
CC marker used in the method of the invention.
XX
SQ Sequence 2813 BP; 868 A; 671 C; 758 G; 516 T; 0 U; 0 Other;

Query Match 23.2%; Score 167.6; DB 13; Length 2813;
Best Local Similarity 98.2%; Pred. No. 7.6e-43;
Matches 167; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 419 AGGTGGTGTGAGTCACCAAGAGAGATGCTGACCTGAAACCCACGAAAGCGAACTCAAAA 478
DB 1298 ATGTGGTGTGAGTCACCAAGAGAGATGCTGACCTGAAACCCACGAAAGCGAACTCAAAA 1357

QY 479 AAACAGTGGACCTCAACGAAGAGAGATGCTGACCTGAAACCCACGAAAGCGAACTCAAAA 538
```

```
|||||
DB 1358 AAACAGTGGACCTCAACGAAGAGATGCTGACCTGAAACCCACGAAAGCGAACTCAAAA 1417
QY 539 GAGAAGAAAGCAAAACCAAGAGAGATGCTGATGCGGCTTCTCAGACAAATG 588
DB 1418 GAGAAGAAAGCAAAACCAAGAGAGATGCTGATGCGGCTTCTCAGACAAATG 1467

RESULT 8
ID AAV09025 standard; cDNA; 3066 BP.
XX
AC AAV09025;
XX
DT 21-JUL-1998 (first entry)
XX
DE Homo sapiens 20q13 amplicon 1b1 transcript.
XX
KW 20q13 amplicon; chromosome 20; tumour; detection;
KW chromosomal abnormalities; probe; gene therapy; antisense inhibition;
KW treatment; age-related macular degeneration; retinitis pigmentation;
KW Leber's congenital amaurosis; ds.
XX
OS Homo sapiens.
XX
FN WO9802539-A1.
XX
PD 22-JAN-1998.
XX
PF 15-JUL-1997; 97WO-US012343.
XX
PR 15-JUL-1996; 96US-00680395.
PR 16-OCT-1996; 96US-00731499.
PR 17-JAN-1997; 97US-00785532.
XX
PA (REGC ) UNIV CALIFORNIA.
XX
PI Gray JW, Collins CC, Hwang S, Godfrey T, Kowbel D, Rommens J;
XX
DR WPI; 1998-110587/10.
XX
PT New sequences from the 20q13 amplicon - used for detecting chromosomal
PT abnormalities, particularly tumours, and for developing products for
PT treating diseases.
XX
PS Claim 1; Page 69-70; 91pp; English.
XX
CC The sequence is that of a cDNA sequence 1b1, which was isolated from the
CC 20q13 amplicon, it is overexpressed in numerous breast cancer cell lines
CC and some primary tumours. It can be used as a probe for the detection of
CC chromosomal abnormalities at 20q13. It and other sequences isolated from
CC the 20q13 amplicon are consistently amplified in primary tumours. These
CC sequences are useful as probes or as probe targets for monitoring the
CC relative copy number of corresponding sequences from a biological sample
CC such as tumour cells. The sequences can also be used in therapeutic
CC applications for modulating the expression of the endogenous gene or the
CC activity of the gene product. Examples of therapeutic approaches include
CC antisense inhibition of gene expression, gene therapy, and monoclonal
CC antibodies that specifically bind the gene products. The products can
CC also be used in the treatment of other diseases, e.g. age-related macular
CC degeneration, Leber's congenital amaurosis and retinitis pigmentation
XX
SQ Sequence 3066 BP; 944 A; 741 C; 818 G; 563 T; 0 U; 0 Other;

Query Match 23.2%; Score 167.6; DB 2; Length 3066;
Best Local Similarity 98.2%; Pred. No. 7.9e-43;
Matches 167; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 419 AGGTGGTGTGAGTCACCAAGAGAGATGCTGACCTGAAACCCACGAAAGCGAACTCAAAA 478
DB 1433 ATGTGGTGTGAGTCACCAAGAGAGATGCTGACCTGAAACCCACGAAAGCGAACTCAAAA 1492

QY 479 AAACAGTGGACCTCAACGAAGAGAGATGCTGACCTGAAACCCACGAAAGCGAACTCAAAA 538
```

```
Db 1493 A A A C G T G A C T C A A C G A A G A G A T G C T G C A C C T G A A C C C A C A A G C G A A A C T C A A A A 1552
QY 539 G A G A A A A A G C A A A C C A A G A A C T C T C T G A T G C G T T T C T C A G A C A A A T G 588
Db 1553 G A G A A A A A G C A A A C C A A G A A C T C T C T G A T G C G T T T C T C A G A C A A A T G 1602

RESULT 9
ADQ23347
ID ADQ23347 standard; DNA; 3418 BP.
XX AC
XX ADQ23347;
XX DT
XX 26-AUG-2004 (first entry)
XX DE
XX Human soft tissue sarcoma-upregulated DNA - SEQ ID 6167.
XX DE
XX soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human;
XX KW ds.
XX OS
XX Homo sapiens.
XX OS
XX WO2004048938-A2.
XX FN
XX 10-JUN-2004.
XX XX
XX 26-NOV-2003; 2003WO-US038193.
XX PF
XX 26-NOV-2002; 2002US-0429739P.
XX PR
XX (PROT-) PROTEIN DESIGN LABS INC.
XX PA
XX Aziz N, Ginsburg WM, Zlotnik A;
XX PI
XX WPI; 2004-441208/41.
XX DR
XX Early detection of soft tissue sarcoma comprises determining expression
XX of a gene in a first soft tissue sample and a normal soft tissue sample
XX PT and comparing the gene expression, also useful in treating soft tissue
XX PT sarcoma.
XX PT
XX Example 2; SEQ ID NO 6167; 210pp; English.
XX PS
XX The invention relates to a novel method for detecting soft tissue sarcoma
XX CC which comprises obtaining a first soft tissue sample from an individual
XX CC and a normal soft tissue sample from the same or different individual,
XX CC determining the expression of a gene in both samples and comparing the
XX CC expression of the gene in both soft tissue samples, where a higher level
XX CC of protein expression in the first soft tissue sample indicates the
XX CC presence of soft tissue sarcoma. The method of the invention has
XX CC cytotstatic applications and may be useful for detecting soft tissue
XX CC sarcoma, possibly via gene therapy or vaccine production. The nucleic
XX CC acid sequences may be useful in diagnostic and screening applications.
XX CC The current sequence is that of a human soft tissue sarcoma-upregulated
XX CC DNA of the invention. The current sequence is not shown within the
XX CC specification per se but was submitted in CD format by the inventor.
XX CC
XX SQ Sequence 3418 BP; 1029 A; 763 C; 842 G; 669 T; 0 U; 115 Other;

Query Match 23.2%; Score 167.6; DB 12; Length 3418;
Best Local Similarity 98.2%; Pred. No. 8.4e-43;
Matches 167; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 419 A G T G T G T G T G A G T C A C C A G T A G A G A T T A A A G T C C A A G G A A G T A G A T C A G C T T A C 478
Db 1346 A T G T G T G T G A G T C A C C A G T A G A T T A A A G T C C A A G G A A G T A G A T C A G C T T A C 1405
QY 479 A A A C A G T G G A C C T C A A C G A A G G A G A T G C T C A C C T G A A C C C A C W G A A G C G A A A C T C A A A A 538
Db 1406 A A A C A G T G G A C C T C A A C G A A G G A G A T G C T C A C C T G A A C C C A C C A A G C G A A A C T C A A A A 1465
QY 539 G A G A A A A A G C A A A C C A A G A A C C T C T C T G A T G C G T T T C T C A G A C A A A T G 588
```

```
Db 1466 G A G A A A A A G C A A A C C A A G A A C C T C T C T G A T G C G T T T C T C A G A C A A A T G 1515

RESULT 10
ACC90605
ID ACC90605 standard; DNA; 2020 BP.
XX AC
XX ACC90605;
XX DT
XX 12-AUG-2003 (first entry)
XX DE
XX Human CGDD-28 encoding DNA.
XX KW
XX Human; cytostatic; antiarteriosclerotic; anticonvulsant; nootropic;
XX neuroprotective; cerebroprotective; anti-HIV; antiallergic;
XX antiinflammatory; gynaecological; cancer; atherosclerosis; epilepsy;
XX Huntington's disease; stroke; AIDS; allergy; placenta; reproductive;
XX CGDD; cell growth; cell differentiation; cell death; gene; ds.
XX OS
XX Homo sapiens.
XX OS
XX WO2003027263-A2.
XX FN
XX 03-APR-2003.
XX PD
XX 26-SEP-2002; 2002WO-US031095.
XX PF
XX 28-SEP-2001; 2001US-0326389P.
XX PR
XX 05-OCT-2001; 2001US-0327380P.
XX PR
XX 05-OCT-2001; 2001US-0328186P.
XX PR
XX 12-OCT-2001; 2001US-0329690P.
XX PR
XX 26-OCT-2001; 2001US-0345384P.
XX PR
XX 26-OCT-2001; 2001US-0348165P.
XX PR
XX 02-NOV-2001; 2001US-0350219P.
XX PR
XX 09-NOV-2001; 2001US-034518P.
XX PR
XX 09-NOV-2001; 2001US-0345143P.
XX PR
XX 16-NOV-2001; 2001US-0332375P.
XX PR
XX 03-DEC-2001; 2001US-0336908P.
XX PR
XX 07-DEC-2001; 2001US-0340747P.
XX XX
XX (INCY-) INCYTE GENOMICS INC.
XX PA
XX Azimzai Y, Baughn MR, Becha SD, Borowsky ML, Chawla NK;
XX PI Elliott VS, Emerling BM, Gandhi AR, Gietzen KJ, Gorvad AE;
XX PI Griffin JA, Hafalia AJA, Ison CH, Kable AE, Kalafus DP;
XX PI Lehr-Mason PM, Lu DAM, Marquis JP, Nguyen DB, Ramkumar J;
XX PI Richardson IW, Sapperstein SK, Swarnakar A, Tang YT, Tran UK;
XX PI Warren BA, Xu Y, Yao MG, Yue H, Yue H;
XX DR
XX WPI; 2003-421159/39.
XX DR P-PSDB; ABR69628.
XX XX
XX New human proteins associated with cell growth, differentiation, and
XX PT death (CGDD), useful for diagnosing, treating and preventing diseases or
XX PT conditions associated with the aberrant CGDD expression e.g. cancer,
XX PT AIDS, or epilepsy.
XX XX
XX Claim 5; Page 330; 350pp; English.
XX PS
XX The invention relates to an isolated polypeptide associated with cell
XX CC growth, differentiation and death (CGDD). Also disclosed are the
XX CC polynucleotides encoding the polypeptides. The polypeptides and
XX CC polynucleotides are useful in diagnosing, treating and preventing
XX CC diseases or conditions associated with the decreased expression or over
XX CC expression of CGDD. Such diseases include cell proliferative (e.g.
XX CC cancer, atherosclerosis), neurological (e.g. epilepsy, Huntington's
XX CC disease, stroke), immune/inflammatory (e.g. AIDS, allergies) and
XX CC reproductive disorders, or disorders of the placenta. They are also
XX CC useful in assessing the effects of exogenous compounds on the expression
XX CC of nucleic acid and amino acid sequences of CGDD. The CGDD or its
XX CC fragments are useful in screening compounds for effectiveness as an
XX CC agonist or antagonist of the polypeptides, or in altering the expression
```

CC of the target polynucleotide and compounds that specifically bind to or
 CC modulate the activity of the polypeptide. Microarrays consisting
 CC polynucleotides of the invention are useful in monitoring or measuring
 CC protein-protein interactions, drug-target interactions, and gene
 CC expression profiles. Sequences given in records ACC90578-ACC90634
 CC represent polynucleotides encoding CGDD proteins of the invention
 XX
 SQ Sequence 2020 BP; 621 A; 525 C; 529 G; 345 T; 0 U; 0 Other;
 Query Match 23.0%; Score 166; DB 8; Length 2020;
 Best Local Similarity 97.6%; Pred. No. 2.1e-42;
 Matches 166; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
 QY 419 AGTGCTGTGAGTCACCACTAGAGATTATAAGTCCAAGAGTAGAATCAGCCTTAC 478
 DB 1312 ATGTGCTGTGAGTCACCACTAGAGATTATAAGTCCAAGAGTAGAATCAGCCTTAC 1371
 QY 479 AAACAGTGGACCTCAACGAAGGAGATGCTGCACCTGAAACCCACACGAAAGCAACTCAAAA 538
 DB 1372 AAACAGTGGACCTCAACGAAGGAGATGCTGCCTGAAACCCACACGAAAGCAACTCAAAA 1431
 QY 539 GAGAGAAAGCAACCAAGAACCTCTCTGATGRCGTTTCTCAGACAAATG 588
 DB 1432 GAGAGAAAGCAACCAAGAACCTCTCTGATGRCGTTTCTCAGACAAATG 1481
 RESULT 11
 ID ABX44866 standard; cDNA; 383 BP.
 XX
 AC ABX44866;
 XX
 DT 21-FEB-2003 (first entry)
 XX
 DE Bovine EST associated with lactation/muscle/fat deposition #10031.
 XX
 KW Bovine; ss; EST; expressed sequence tag; lactation; LMFD;
 KW muscle deposition; fat deposition; genome mapping; gene identification;
 KW gene analysis; cattle breeding.
 XX
 OS Bos Taurus.
 XX
 PN US2002137139-A1.
 XX
 PD 26-SEP-2002.
 XX
 PF 24-SEP-2001; 2001US-00960352.
 XX
 PR 12-JAN-1999; 99US-0115707P.
 PR 11-JAN-2000; 2000US-00480902.
 XX
 PA (BYATT/) BYATT J C.
 PA (MATH/) MATHIALAGAN N.
 PA (TAON/) TAO N.
 PA (WARR/) WARREN W C.
 XX
 PI Byatt JC, Mathialagan N, Tao N, Warren WC;
 XX
 DR WPI; 2003-110599/10.
 XX
 PT New nucleic acid associated with lactation, and muscle and fat
 PT deposition, useful for genome mapping, gene identification and analysis,
 PT cattle breeding, or for genetically improving cattle.
 XX
 PS Claim 2; SEQ ID NO 10031; 245pp; English.
 XX
 CC The invention relates to a purified nucleic acid molecule associated with
 CC lactation or muscle and fat deposition (designated LMFD), derived from
 CC cattle, and the LMFD nucleic acid can specifically hybridise to a second
 CC nucleic acid molecule comprising any of 15112 nucleotide sequences,
 CC appearing as ABX34836-ABX49947, or complements of them. Also included are
 CC ; (1) a transformed cell having a nucleic acid comprising an LMFD nucleic
 CC acid linked to a promoter and a 3' non- translated sequence that

CC functions in the cell to cause termination of transcription and addition
 CC of polyadenylated ribonucleotides to a 3' end of the mRNA molecule; and
 CC (2) determining a level or pattern of a molecule in a bovine cell or
 CC tissue comprising: (a) incubating a marker nucleic acid (comprising any
 CC of the 15112 nucleic acid sequences or its complement or fragment) with a
 CC complementary nucleic acid molecule obtained from the bovine cell or
 CC tissue, where hybridisation between the marker nucleic acid and the
 CC complementary nucleic acid permits the detection of the molecule; and (b)
 CC detecting the level or pattern of the complementary nucleic acid, where
 CC the detection of the complementary nucleic acid is predictive of the
 CC level or pattern of the molecule. The LMFD nucleic acid is used for
 CC determining a level or pattern of a molecule in a bovine cell or tissue.
 CC It is useful for genome mapping, gene identification and analysis, cattle
 CC breeding, preparation of constructs for use in cattle gene expression, or
 CC for genetically improving cattle. The present sequence is one of the
 CC 15112 bovine LMFD EST (expressed sequence tag) nucleic acids. Note: The
 CC present sequence was not shown in the specification but was obtained in
 CC electronic format from the USPTO web site:
 CC seqdata.uspto.gov/sequence.html?DocID=20020137139
 XX
 SQ Sequence 383 BP; 123 A; 95 C; 108 G; 56 T; 0 U; 1 Other;
 Query Match 12.3%; Score 88.6; DB 8; Length 383;
 Best Local Similarity 62.9%; Pred. No. 9.8e-18;
 Matches 168; Conservative 2; Mismatches 91; Indels 6; Gaps 2;
 QY 328 GGACAGGGGGGCGACGAAGGACCCCACTTTGTAAACATGTTTGGGGGAATGTTTGT 387
 DB 51 GAAAAAGGGGGGGCCCGAGCTCTCTGCTCTGGCAAACTGTTTGGAAAAAGTCAGTTAA 110
 QY 388 TTCATTTTCTATTACCTGGCAAAATATCCAGTGGTGTGTGAGTCCAGTAGAGATT 447
 DB 111 GAGGATTTCAGTCCCCACAGGTGCAGAGGAGAAATGTTGTGTGAGTCCAGTAGAGCT 170
 QY 448 ATAAAGTCCAGGAAGTAGAATCAGCCTTACAAACAGTGGACCT---CAACGAAGGAGAT 504
 DB 171 GTAAAGTCTGAAGAAGTAGAGTACGCTTACAAACAGTGGATCTCGGCGAAGAGGAGAG 230
 QY 505 GCTGCACCTGAACCCACACGAGCGAAACTCAAAAGAGAAGAAAGCAAA---CCAAGAAC 561
 DB 231 CGCACACCCGATCCCGCAGAAGAGCACCACGAGAGAGAAGAACGAGCCCGAGGCC 290
 QY 562 TCTCTGATGRCGTTTCTCAGACAAATG 588
 DB 291 TCCCTGATGGCGTCTCTCAGACAGATG 317
 RESULT 12
 ID ADQ53268 standard; DNA; 540 BP.
 XX
 AC ADQ53268;
 XX
 DT 21-OCT-2004 (first entry)
 XX
 DE Novel canine microarray-related DNA sequence SeqID4570.
 XX
 KW canine microarray; drug screening; toxicity assay;
 KW environmental pollutant; cellular response; gene expression profile;
 KW toxic response; liver necrosis; fatty liver disease;
 KW protein adduct formation; hepatitis; dog; ds.
 XX
 OS Canis familiaris.
 XX
 PN WO2004063324-A2.
 XX
 PD 29-JUL-2004.
 XX
 PF 05-MAY-2003; 2003WO-US013853.
 XX
 PR 03-MAY-2002; 2002US-0377240P.
 XX
 PA (GENE-) GENE LOGIC INC.

PA (PFIZ) PFIZER PROD INC.
 XX Diggins JC, Porter M, Wei T;
 XX WPI; 2004-561890/54.
 DR
 XX New isolated nucleic acid molecule, useful for drug screening and
 PT toxicity assays or for assessing the impact, including toxicity, of a
 PT compound, pharmaceutical agent or environmental pollutant on a cell or
 PT living organism.
 XX
 PS Claim 1; SEQ ID NO 4570; 41pp; English.
 XX
 CC This invention is related to a novel isolated canine nucleic acid
 CC sequences and the construction of canine microarrays containing a
 CC significant portion of the canine genome. The isolated canine nucleic
 CC acid sequences of the invention may be useful for drug screening and
 CC toxicity assays. The invention is therefore useful for assessing the
 CC impact, including toxicity, of a compound, pharmaceutical agent or
 CC environmental pollutant on a cell or living organism. The methods are
 CC useful for detecting genes that are up- or down-regulated in canines in a
 CC disease state. The sequences are useful as diagnostic agents or markers
 CC to detect a cellular response in a sample individually or as part of a
 CC gene expression profile. It is also useful as a target for agents that
 CC modulate gene expression or activity. The database is useful for
 CC producing electronic Northernblots that allow the user to determine the cell
 CC type or tissue in which a given gene is expressed and to allow
 CC determination of the abundance or expression level of a given gene in a
 CC particular tissue or cell. The methods are useful for determining the
 CC similarity of a toxic response to one or more individual compounds. The
 CC methods are useful for predicting at least one toxic response or the
 CC likelihood that a compound or test agent will induce various specific
 CC pathologies such as those of the liver (liver necrosis, fatty liver
 CC disease, protein adduct formation or hepatitis), those of the kidney,
 CC heart, brain or testes, or other pathologies associated with at least one
 CC of the toxins. The methods are also useful for predicting or elucidating
 CC the potential cellular pathways influenced, induced or modulated by the
 CC compound or test agent due to the similarity of the expression profile
 CC compared to the profile induced by a known toxin. The present sequence is
 CC that of a canine DNA sequence which was claimed for use during the
 CC production of a canine microarray of the invention.
 XX
 SQ Sequence 540 BP; 167 A; 137 C; 140 G; 74 T; 0 U; 22 Other;

Query Match 9.7%; Score 70.2; DB 13; Length 540;
 Best Local Similarity 80.2%; Pred. No. 1.1e-11;
 Matches 101; Conservative 2; Mismatches 17; Indels 6; Gaps 2;
 QY 469 TCAGCCTTACAAACAGTGGACCTCA---ACGAGGAGATGCTGCACCTGAACCCACWGA 525
 DB 1 TCAGNNNTACAAACAGTNGATCTCAATGAAGAGGAGATGCCACANNTGAACCCACAGAN 60
 QY 526 GCGAAACTCAAAAGAGAAGAACCAACCA---AGAACCTCTCTGATGCGTTTCTCAGA 582
 DB 61 GTAAAACTCAAAAGAGAAGAAATACAAACCCAGGAGAACCTCCCTGATGCGTTTCTCAGA 120
 QY 583 CAAATG 588
 DB 121 CAAATG 126

RESULT 13
 AB213040
 ID AB213040 standard; DNA; 957 BP.
 XX
 AC AB213040;
 XX
 DT 21-JAN-2003 (first entry)
 XX
 DE Arabidopsis thaliana stress regulated gene SEQ ID NO 845.
 XX
 KW Arabidopsis thaliana; plant; gene; stress; transgenic; ds.
 XX

OS Arabidopsis thaliana.
 XX
 PN WO200216655-A2.
 XX
 PD 28-FEB-2002.
 XX
 PF 24-AUG-2001; 2001WO-US026685.
 XX
 PR 24-AUG-2000; 2000US-0227866P.
 PR 26-JAN-2001; 2001US-0264647P.
 PR 22-JUN-2001; 2001US-0300111P.
 XX
 PA (SCRI) SCRIPES RES INST.
 PA (SYGN) SYNGENTA PARTICIPATIONS AG.
 XX
 PI Harper JF, Kreps J, Wang X, Zhu T;
 XX
 DR WPI; 2002-304127/34.
 XX
 PT Identifying a stress condition to which a plant cell has been exposed and
 PT producing plants with increased tolerance to these abiotic stresses.
 XX
 PS Claim 144; SEQ ID NO 845; 577pp + Sequence Listing; English.
 XX
 CC The invention relates to identifying a stress condition to which a plant
 CC cell has been exposed, comprising: (a) contacting nucleic acid
 CC representative of expressed polynucleotides in the plant cell with an
 CC array or probes representative of the plant cell genome; and (b)
 CC detecting a profile of expressed polynucleotides in the plant cell
 CC characteristic of a stress response. The method is useful in the
 CC production of transgenic plants, cells and seeds and in producing plants
 CC with increased tolerance to abiotic stress. The present sequence is that
 CC of an Arabidopsis thaliana stress regulated gene (AB212196-AB217574) used
 CC in methods of the invention. Note: The sequence data for this patent is
 CC not represented in the printed specification but is based on sequence
 CC information supplied to Derwent by the European Patent Office
 XX

SQ Sequence 957 BP; 259 A; 188 C; 237 G; 273 T; 0 U; 0 Other;
 Query Match 5.7%; Score 41.2; DB 6; Length 957;
 Best Local Similarity 47.2%; Pred. No. 0.035;
 Matches 118; Conservative 2; Mismatches 130; Indels 0; Gaps 0;
 QY 382 TTGTTTTCATTTCTTATTACCTGGCAAAATAATCCAGGTGGTGTGTGAGTCACCACTA 441
 DB 358 TGGATGTTTAGAGTCTTTGGACATGACAAAGATGGGTACTTTGATGGAGGCTTGCCGAAA 417
 QY 442 GAGATTATTAAGTCCAGGAAGTAGAATCAGCCTTACAAACAGTGGACCTCAACGAAGA 501
 DB 418 TGGCGTGTCTCAGGCTATGACGTTGAATCCAGTGTTCAAATGATGCCATTTTGAAGGC 477
 QY 502 GATGTGCACTTGAACCCACWGAAGCGAAACTCAAAGAGAGAAGAAAGCAACCAAGAAC 561
 DB 478 AGTGACGCACTGAGGCTATAGAGAAGATCTATCAGGACCAACAATTAGCCCACTA 537
 QY 562 TCTCTGATGRCGTCTTTCAGACAAATGGTAAGCCCTTACTTCCAGTATAGAAACCTAA 621
 DB 538 TTTTCAGACGAAGTTCGGGCCACATCTTGTATTGGCACTTGTACAGTGAAGGAACATC 597
 QY 622 GATACCTAGA 631
 DB 598 GAGGATAAGA 607

RESULT 14
 AAC50675
 ID AAC50675 standard; DNA; 1347 BP.
 XX
 AC AAC50675;
 XX
 DT 18-OCT-2000 (first entry)
 XX
 DE Arabidopsis thaliana DNA fragment SEQ ID NO: 65724.

XX Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway; metabolic pathway;
KW promoter; termination sequence; ss.
XX Arabidopsis thaliana.
OS
XX
PN EP1033405-A2.
XX
PD
XX
PF
XX
PP
XX
XX
PR 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 29-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 23-APR-1999; 99US-0130891P.
PR 28-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137503P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 24-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.

PR	31-AUG-1999;	99US-0151438P.	RESULT 15	
PR	01-SEP-1999;	99US-0151930P.	AAC33342	
PR	07-SEP-1999;	99US-0152363P.	ID AAC33342 standard; DNA; 1352 BP.	
PR	10-SEP-1999;	99US-0153070P.	XX	
PR	13-SEP-1999;	99US-0153758P.	AC AAC33342;	
PR	15-SEP-1999;	99US-0154018P.	XX	
PR	16-SEP-1999;	99US-0154039P.	DT 17-OCT-2000 (first entry)	
PR	20-SEP-1999;	99US-0154779P.	XX	
PR	22-SEP-1999;	99US-0155139P.	XX	Arabidopsis thaliana DNA fragment SEQ ID NO: 2682.
PR	23-SEP-1999;	99US-0155486P.	XX	
PR	24-SEP-1999;	99US-0155659P.	XX	Hybridisation assay; genetic mapping; gene expression control;
PR	28-SEP-1999;	99US-0156458P.	KW	protein identification; signal transduction pathway; metabolic pathway;
PR	29-SEP-1999;	99US-0156596P.	KW	promoter; termination sequence; ss.
PR	04-OCT-1999;	99US-0157117P.	XX	
PR	05-OCT-1999;	99US-0157753P.	XX	
PR	06-OCT-1999;	99US-0157865P.	OS	Arabidopsis thaliana.
PR	07-OCT-1999;	99US-0158029P.	XX	
PR	08-OCT-1999;	99US-0158232P.	PN	EP1033405-A2.
PR	12-OCT-1999;	99US-0158369P.	XX	
PR	13-OCT-1999;	99US-0159293P.	XX	06-SEP-2000.
PR	13-OCT-1999;	99US-0159294P.	XX	
PR	13-OCT-1999;	99US-0159329P.	XX	25-FEB-2000; 2000EP-00301439.
PR	14-OCT-1999;	99US-0159330P.	XX	
PR	14-OCT-1999;	99US-0159331P.	PR	25-FEB-1999; 99US-0121825P.
PR	14-OCT-1999;	99US-0159637P.	PR	05-MAR-1999; 99US-0123180P.
PR	14-OCT-1999;	99US-0159638P.	PR	09-MAR-1999; 99US-0123548P.
PR	18-OCT-1999;	99US-0159584P.	PR	23-MAR-1999; 99US-0125788P.
PR	21-OCT-1999;	99US-0160741P.	PR	25-MAR-1999; 99US-0126264P.
PR	21-OCT-1999;	99US-0160767P.	PR	29-MAR-1999; 99US-0126785P.
PR	21-OCT-1999;	99US-0160770P.	PR	01-APR-1999; 99US-0127462P.
PR	21-OCT-1999;	99US-0160814P.	PR	06-APR-1999; 99US-0128234P.
PR	21-OCT-1999;	99US-0160815P.	PR	08-APR-1999; 99US-0128714P.
PR	22-OCT-1999;	99US-0160981P.	PR	16-APR-1999; 99US-0129845P.
PR	22-OCT-1999;	99US-0160989P.	PR	19-APR-1999; 99US-0130077P.
PR	25-OCT-1999;	99US-0161404P.	PR	21-APR-1999; 99US-0130449P.
PR	25-OCT-1999;	99US-0161405P.	PR	23-APR-1999; 99US-0130510P.
PR	25-OCT-1999;	99US-0161406P.	PR	23-APR-1999; 99US-0130891P.
PR	26-OCT-1999;	99US-0161359P.	PR	28-APR-1999; 99US-0131449P.
PR	26-OCT-1999;	99US-0161360P.	PR	30-APR-1999; 99US-0132048P.
PR	26-OCT-1999;	99US-0161361P.	PR	30-APR-1999; 99US-0132407P.
PR	28-OCT-1999;	99US-0161920P.	PR	04-MAY-1999; 99US-0132484P.
PR	28-OCT-1999;	99US-0161922P.	PR	05-MAY-1999; 99US-0132485P.
PR	28-OCT-1999;	99US-0161933P.	PR	06-MAY-1999; 99US-0132486P.
PR	29-OCT-1999;	99US-0162142P.	PR	06-MAY-1999; 99US-0132487P.
PR			PR	07-MAY-1999; 99US-0132863P.
PR			PR	11-MAY-1999; 99US-0134256P.
PR			PR	14-MAY-1999; 99US-0134218P.
PR			PR	14-MAY-1999; 99US-0134219P.
PR			PR	14-MAY-1999; 99US-0134221P.
PR			PR	14-MAY-1999; 99US-0134370P.
PR			PR	18-MAY-1999; 99US-0134768P.
PR			PR	19-MAY-1999; 99US-0134941P.
PR			PR	20-MAY-1999; 99US-0135124P.
PR			PR	21-MAY-1999; 99US-0135353P.
PR			PR	24-MAY-1999; 99US-0135629P.
PR			PR	25-MAY-1999; 99US-0136021P.
PR			PR	27-MAY-1999; 99US-0136392P.
PR			PR	28-MAY-1999; 99US-0136782P.
PR			PR	01-JUN-1999; 99US-0137222P.
PR			PR	03-JUN-1999; 99US-0137528P.
PR			PR	04-JUN-1999; 99US-0137502P.
PR			PR	07-JUN-1999; 99US-0137724P.
PR			PR	08-JUN-1999; 99US-0138094P.
PR			PR	10-JUN-1999; 99US-0138540P.
PR			PR	10-JUN-1999; 99US-0138847P.
PR			PR	14-JUN-1999; 99US-0139119P.
PR			PR	16-JUN-1999; 99US-0139452P.
PR			PR	16-JUN-1999; 99US-0139453P.
PR			PR	17-JUN-1999; 99US-0139492P.
PR			PR	18-JUN-1999; 99US-0139454P.
PR			PR	18-JUN-1999; 99US-0139455P.
PR			PR	18-JUN-1999; 99US-0139456P.
PR			PR	18-JUN-1999; 99US-0139457P.
PR			PR	18-JUN-1999; 99US-0139458P.

Query Match		5.7%;	Score 41.2;	DB 3;	Length 1347;
Best Local Similarity		47.2%;	Pred. No. 0.042;		
Matches 118;		Conservative 2;	Mismatches 130;	Indels 0;	Gaps 0;
QY	382	TTGTTTTCATTTTCTTATTTACCTGGCAAAATATCCAGGTGGTGTGTGAGTCACCAAGTA	441		
Db	594	TGGATGTTTAGAGTCTTTGGACATGACAAAGTATGGGTACTTGTGAGGCTTGCCGAAA	653		
QY	442	GACATTATAAGTCCAGGAAGTAGAATCAGCCTTACAAACGTGACCTCAACGAAGGA	501		
Db	654	TGGCGTCTTCAGGCTATGACGTGTAATCCAGTGTTCATATGATGCAATTTTGAAGCC	713		
QY	502	GATGCTGCACCTGAACCCACGAGGCAAACTCAAAAGAGAAAGCAACCAAGAAC	561		
Db	714	AGTGCAGCACTGAGGCTATAGAGAGATCTATCAAGGACAAACAATTAGCCCAATACT	773		
QY	562	TCTCTGATGRCGTTTCTCAGACAAATGGTAAGCCCTTACTTCCAGTATAGGAAACCTAA	621		
Db	774	TTTCAGCAAGATTCGCGCCACATCTTGATTGGCATTGTATCGAGGTGAGGAAACATC	833		
QY	622	GATACCTTAGA	631		
Db	834	GAGGATAAGA	843		

```
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 24-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144684P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 16-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.

PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 07-SEP-1999; 99US-0151930P.
PR 10-SEP-1999; 99US-0152363P.
PR 13-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161922P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 5.7%; Score 41.2; DB 3; Length 1352;
Best Local Similarity 47.2%; Pred. No. 0.042;
Matches 118; Conservative 2; Mismatches 130; Indels 0; Gaps 0;

Qy 382 TTGTTTTCATTTTCTTATTACCTGGCAAAATAATCCAGGTGGTGTGTGAGTCCACGTA 441
Db 596 TGGATGTTTAGAGTCTTTGGACATGACAAAGTATGGGTACTTGTATGGAGGCTTGCCGAA 655

Qy 442 GAGATTATAAGTCCAGGAAGTAGATCAGCCTTACAAACAGTGGACCTCAACGAAGA 501
Db 656 TGGCGTCTTTCAGGCTATGACGTTGATCCAGTGTTCATATGATGCAATTTTGAAGCC 715

Qy 502 GATGCTCACCTGAACCCACGAGGCGAAACTCAAAAGAGAGAAAGCAACCAAGAAC 561
Db 716 AGTCGCAACTGAGGCTATAGAGAAGATCTATCAGGACAAACATTTAGCCCAATAACT 775
```

QY	562	TCTCTGATGRCGTTTCTCAGACAAATGGTAAGCCCTTACTTCCAGTATAGGAAACCTAA	621
Db	776	TTTCAGACGAAGTTCGGCCACATCTTGTATTGGCACTTGATCAGGTGAAGGAAACATC	835
QY	622	GATACCTAGA	631
Db	836	GAGGATAGA	845

Search completed: May 5, 2005, 00:37:07
Job time : 507.471 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 4, 2005, 23:46:20 ; Search time 162.426 Seconds
(without alignments)
7283.499 Million cell updates/sec

Title: US-08-731-499-2

Perfect score: 723

Sequence: 1 TGGAGCTGTCATGTTACC.....TGAATGGATATGGATTGCG 723

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA.*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PTUS_COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	722.2	99.9	723	2	US-08-680-395-2
2	722.2	99.9	723	4	US-08-892-695-2
3	714.8	98.9	130563	4	US-09-949-016-12273
4	714.8	98.9	131379	4	US-09-949-016-16050
5	167.6	23.2	2813	3	US-09-689-255C-3
6	167.6	23.2	2813	4	US-09-949-016-531
7	167.6	23.2	3066	4	US-08-892-695-12
8	167.6	23.2	3629	4	US-09-949-016-4308
9	41.8	5.8	7218	1	US-08-232-463-14
10	36	5.0	1728	4	US-09-248-796A-3540
11	34.8	4.8	601	4	US-09-949-016-162024
12	34.8	4.8	601	4	US-09-949-016-204212
13	34.8	4.8	9668	4	US-09-949-016-4532
14	34.8	4.8	28374	4	US-09-949-016-17508
15	34.8	4.8	238815	4	US-09-949-016-16274
16	34.4	4.8	415	4	US-09-270-767-8451
17	34.4	4.8	415	4	US-09-270-767-23733
18	34.4	4.8	601	4	US-09-949-016-204213
19	34.4	4.8	68035	4	US-09-949-016-16219
20	34.2	4.7	832	4	US-09-621-976-2813
21	34.2	4.7	45323	4	US-09-949-016-16142
22	34	4.7	3183	4	US-09-248-796A-2521
23	34	4.7	74353	4	US-09-949-016-15336
24	33.8	4.7	597	4	US-09-270-767-5349
25	33.8	4.7	597	4	US-09-270-767-20631
26	33.4	4.6	601	4	US-09-949-016-81376
27	33.4	4.6	1596	4	US-09-248-796A-1457

28	33.4	4.6	121982	4	US-09-949-016-12085	Sequence 12085, A
29	33.4	4.6	121982	4	US-09-949-016-14105	Sequence 14105, A
30	33.2	4.6	1584	1	US-07-667-276A-1	Sequence 1, Appli
31	33.2	4.6	2709	4	US-10-087-402-5	Sequence 5, Appli
32	33.2	4.6	2746	2	US-08-576-165-3	Sequence 3, Appli
33	33	4.6	371	3	US-08-943-731-103	Sequence 103, App
34	33	4.6	534	4	US-09-387-286-26	Sequence 26, Appli
35	33	4.6	558	4	US-09-270-767-7078	Sequence 7078, Ap
36	33	4.6	558	4	US-09-270-767-22360	Sequence 22360, A
37	33	4.6	601	4	US-09-949-016-37900	Sequence 37900, A
38	33	4.6	601	4	US-09-949-016-176986	Sequence 176986,
39	33	4.6	1236	4	US-09-540-236-1364	Sequence 1364, Ap
40	33	4.6	1989	4	US-09-248-796A-1350	Sequence 1350, Ap
41	33	4.6	36682	3	US-08-943-731-2	Sequence 2, Appli
42	33	4.6	40315	4	US-09-949-016-11753	Sequence 11753, A
43	33	4.6	40649	4	US-09-949-016-14219	Sequence 14219, A
44	33	4.6	41062	4	US-09-949-016-12761	Sequence 12761, A
45	33	4.6	41062	4	US-09-949-016-16796	Sequence 16796, A

ALIGNMENTS

RESULT 1
US-08-680-395-2
; Sequence 2, Application US/08680395
; Patent No. 5892010
; GENERAL INFORMATION:
; APPLICANT: Gray, Joe W.
; APPLICANT: Colling, Colin
; APPLICANT: Hwang, Soo-in
; APPLICANT: Godfrey, Tony
; APPLICANT: Kowbel, David
; APPLICANT: Rommens, Johanna
; TITLE OF INVENTION: Genes from the 20q13 Amplicon and Their
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/680,395
; FILING DATE: 15-JUL-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Bastian, Kevin L.
; REGISTRATION NUMBER: 34,774
; REFERENCE/DOCKET NUMBER: 023070-0689000US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 723 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; FEATURE:
; NAME/KEY: -
; LOCATION: 1..723
; OTHER INFORMATION: /note= "cdna clone 1b11 of 3.5kb
; OTHER INFORMATION: transcript"

US-08-680-395-2


```
RESULT 3
US-09-949-016-12273
; Sequence 12273, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12273
; LENGTH: 130563
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-12273

Query Match      98.9%; Score 714.8; DB 4; Length 130563;
Best Local Similarity 99.2%; Pred. No. 1.8e-224;
Matches 716; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 1 TGGAGCTGTCATGTTACCGTCTCTAACGTTGGACTCTTAAGAAATGATTATTCCTGG 60
Db 105047 TGGAGCTGTCATGTTACCGTCTCTAACGTTGGACTCTTAAGAAATGATTATTCCTGG 105106

Qy 61 TTCTAGACAGCCCAAAATGTAATTCACCTACGTGCAGATTAAAGAGTGGCTTACTAG 120
Db 105107 TTCTAGACAGCCCAAAATGTAATTCACCTACGTGCAGATTAAAGAGTGGCTTACTAG 105166

Qy 121 ATTTGATTGGGTATTGAGCATGCTCTGAATGACAGTCCCCAAAAAGGACCTCTTATCCGT 180
Db 105167 ATTTGATTGGGTATTGAGCATGCTCTGAATGACAGTCCCCAAAAAGGACCTCTTATCCGT 105226

Qy 181 TCTTCCCTTGGGGAAGGCTTTTGGCAGCTTCCATGTCATGTCAGTGGCAGTTGAGCTTGAA 240
Db 105227 TCTTCCCTTGGGGAAGGCTTTTGGCAGCTTCCATGTCATGTCAGTGGCAGTTGAGCTTGAA 105286

Qy 241 ATTTGATGGTGTACCAACATAGCAATTAATTCACCTGTCAGTGGCAGTTGAGCTTGAA 300
Db 105287 ATTTGATGGTGTACCAACATAGCAATTAATTCACCTGTCAGTGGCAGTTGAGCTTGAA 105346

Qy 301 ATAGATTCAAACCTGACAGTACAGATTATAAGTCCAAAGGAGTAGAATCAGCTTACAA 480
Db 105347 ATAGATTCAAACCTGACAGTACAGATTATAAGTCCAAAGGAGTAGAATCAGCTTACAA 105406

Qy 361 AACATGTTTGGGGAATGTTTGTGTTTTCATTTTCTTATTCCTGGCAAAATATCCAG 420
Db 105407 AACATGTTTGGGGAATGTTTGTGTTTTCATTTTCTTATTCCTGGCAAAATATCCAG 105586

Qy 541 GAAGAAGCAAAACCAAGAACTCTCTGATGTCGTTTCTCAGACAAATGGTAAGCCCTTA 600
Db 105587 GAAGAAGCAAAACCAAGAACTCTCTGATGTCGTTTCTCAGACAAATGGTAAGCCCTTA 105646

Qy 601 CTTCCAGTATAGGAACCTTAAGATACCTAGACGGCTTTTGGGAACAATGGCTCATGCC 660
Db 105647 CTTCCAGTATAGGAACCTTAAGATACCTAGACGGCTTTTGGGAACAATGGCTCATGCC 105706
```

```
Qy 661 ACAGTAGTAGGACACATAATTGTAGCTGCTGTCTATGGAATGTGAATGGAATATGGATT 720
Db 105707 ACAGTAGTAGGACACATAATTGTAGCTGCTGTCTATGGAATGTGAATGGAATATGGATT 105766

Qy 721 GC 722
Db 105767 GC 105768

RESULT 4
US-09-949-016-16050
; Sequence 16050, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16050
; LENGTH: 131379
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-16050

Query Match      98.9%; Score 714.8; DB 4; Length 131379;
Best Local Similarity 99.2%; Pred. No. 1.8e-224;
Matches 716; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 1 TGGAGCTGTCATGTTACCGTCTCTAACGTTGGACTCTTAAGAAATGATTATTCCTGG 60
Db 105047 TGGAGCTGTCATGTTACCGTCTCTAACGTTGGACTCTTAAGAAATGATTATTCCTGG 105106

Qy 61 TTCTAGACAGCCCAAAATGTAATTCACCTACGTGCAGATTAAAGAGTGGCTTACTAG 120
Db 105107 TTCTAGACAGCCCAAAATGTAATTCACCTACGTGCAGATTAAAGAGTGGCTTACTAG 105166

Qy 121 ATTTGATTGGGTATTGAGCATGCTCTGAATGACAGTCCCCAAAAAGGACCTCTTATCCGT 180
Db 105167 ATTTGATTGGGTATTGAGCATGCTCTGAATGACAGTCCCCAAAAAGGACCTCTTATCCGT 105226

Qy 181 TCTTCCCTTGGGGAAGGCTTTTGGCAGCTTCCATGTCATGTCAGTGGCAGTTGAGCTTGAA 240
Db 105227 TCTTCCCTTGGGGAAGGCTTTTGGCAGCTTCCATGTCATGTCAGTGGCAGTTGAGCTTGAA 105286

Qy 241 ATTTGATGGTGTACCAACATAGCAATTAATTCACCTGTCAGTGGCAGTTGAGCTTGAA 300
Db 105287 ATTTGATGGTGTACCAACATAGCAATTAATTCACCTGTCAGTGGCAGTTGAGCTTGAA 105346

Qy 301 ATAGATTCAAACCTGACAGTACAGATTATAAGTCCAAAGGAGTAGAATCAGCTTACAA 360
Db 105347 ATAGATTCAAACCTGACAGTACAGATTATAAGTCCAAAGGAGTAGAATCAGCTTACAA 105406

Qy 361 AACATGTTTGGGGAATGTTTGTGTTTTCATTTTCTTATTCCTGGCAAAATATCCAG 420
Db 105407 AACATGTTTGGGGAATGTTTGTGTTTTCATTTTCTTATTCCTGGCAAAATATCCAG 105466

Qy 421 GTGCTGTGTAGTCAACAGTAGAGATTATAAGTCCAAAGGAGTAGAATCAGCTTACAA 480
Db 105467 GTGCTGTGTAGTCAACAGTAGAGATTATAAGTCCAAAGGAGTAGAATCAGCTTACAA 105526

Qy 481 ACAGTGGACCTCAACGAGGAGTAGCTGCACCTGACCTGACCCAGGAGCCCACTTTGTA 540
Db 105527 ACAGTGGACCTCAACGAGGAGTAGCTGCACCTGACCTGACCCAGGAGCCCACTTTGTA 105586

Qy 541 GAAGAAGCAAAACCAAGAACTCTCTGATGTCGTTTCTCAGACAAATGGTAAGCCCTTA 600
Db 105587 GAAGAAGCAAAACCAAGAACTCTCTGATGTCGTTTCTCAGACAAATGGTAAGCCCTTA 105646

Qy 601 CTTCCAGTATAGGAACCTTAAGATACCTAGACGGCTTTTGGGAACAATGGCTCATGCC 660
Db 105647 CTTCCAGTATAGGAACCTTAAGATACCTAGACGGCTTTTGGGAACAATGGCTCATGCC 105706
```

```
QY 541 GAAGAAAGCAAAACCAAGAACCTCTCTGATGRCGTTTCTCAGACAAATGGTAAGCCCTTAA 600
|||
Db 105587 GAAGAAAGCAAAACCAAGAACCTCTCTGATGCGGTTTCTCAGACAAATGGTAAGCCCTTAA 105646
|||
QY 601 CTTCCAGTATAGGAACCTTAAGATACCTAGAGCGGCTTTTGGAAACAATGGGCTCATGCC 660
|||
Db 105647 CTTCCAGTATAGGAACCTTAAGATACCTAGAGCGGCTTTTGGAAACAATGGGCTCATGCC 105706
|||
QY 661 ACAGGTAGTAGGACACATAATTCTAGCTGTGTATGGAATCTGAATGGAATATCGATT 720
|||
Db 105707 ACAGGTAGTAGGACACATAATTCTAGCTGTGTATGGAATCTGAATGGAATATCGATT 105766
|||
QY 721 GC 722
|||
Db 105767 GC 105768

RESULT 5
US-09-689-255C-3
; Sequence 3, Application US/09689255C
; Patent No. 6395544
; GENERAL INFORMATION:
; APPLICANT: Lex M. Cowser
; APPLICANT: Susan M. Freier
; TITLE OF INVENTION: ANTISENSE MODULATION OF BCAS1 EXPRESSION
; FILE REFERENCE: RTS-0171
; CURRENT APPLICATION NUMBER: US/09/689,255C
; CURRENT FILING DATE: 2000-10-11
; NUMBER OF SEQ ID NOS: 47
; SEQ ID NO 3
; LENGTH: 2813
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (118)...(1872)
US-09-689-255C-3

Query Match 23.2%; Score 167.6; DB 3; Length 2813;
Best Local Similarity 98.2%; Pred. No. 1.8e-44;
Matches 167; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 419 AGTGTGTGTGTGAGTCACCAGTAGAGATTATAAAGTCCAAGGAAGTAGAATCAGCCTTAC 478
|||
Db 1298 ATGTGTGTGTGAGTCACCAGTAGAGATTATAAAGTCCAAGGAAGTAGAATCAGCCTTAC 1357
|||
QY 479 AAACAGTGGACCTCAACGAAGGAGATGCTGCACCTTGAACCCACGAAAGCAAACTCAAAA 538
|||
Db 1358 AAACAGTGGACCTCAACGAAGGAGATGCTGCACCTTGAACCCACGAAAGCAAACTCAAAA 1417
|||
QY 539 GAGAAGAAAGCAAAACCAAGAACCTCTCTGATGRCGTTTCTCAGACAAATG 588
|||
Db 1418 GAGAAGAAAGCAAAACCAAGAACCTCTCTGATGRCGTTTCTCAGACAAATG 1467
|||

RESULT 6
US-09-949-016-531
; Sequence 531, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08

Query Match 23.2%; Score 167.6; DB 4; Length 3066;
Best Local Similarity 98.2%; Pred. No. 1.9e-44;
Matches 167; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 419 AGTGTGTGTGTGAGTCACCAGTAGAGATTATAAAGTCCAAGGAAGTAGAATCAGCCTTAC 478
|||
Db 1433 ATGTGTGTGTGAGTCACCAGTAGAGATTATAAAGTCCAAGGAAGTAGAATCAGCCTTAC 1492
|||
QY 479 AAACAGTGGACCTCAACGAAGGAGATGCTGCACCTTGAACCCACGAAAGCAAACTCAAAA 538
|||
Db 1493 AAACAGTGGACCTCAACGAAGGAGATGCTGCACCTTGAACCCACGAAAGCAAACTCAAAA 1552
|||
QY 539 GAGAAGAAAGCAAAACCAAGAACCTCTCTGATGRCGTTTCTCAGACAAATG 588
|||
Db 1553 GAGAAGAAAGCAAAACCAAGAACCTCTCTGATGRCGTTTCTCAGACAAATG 1602
|||
```

```
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 531
; LENGTH: 2813
; TYPE: DNA
; ORGANISM: Human
; ORGANISM: Human
US-09-949-016-531

Query Match 23.2%; Score 167.6; DB 4; Length 2813;
Best Local Similarity 98.2%; Pred. No. 1.8e-44;
Matches 167; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 419 AGTGTGTGTGTGAGTCACCAGTAGAGATTATAAAGTCCAAGGAAGTAGAATCAGCCTTAC 478
|||
Db 1298 ATGTGTGTGTGAGTCACCAGTAGAGATTATAAAGTCCAAGGAAGTAGAATCAGCCTTAC 1357
|||
QY 479 AAACAGTGGACCTCAACGAAGGAGATGCTGCACCTTGAACCCACGAAAGCAAACTCAAAA 538
|||
Db 1358 AAACAGTGGACCTCAACGAAGGAGATGCTGCACCTTGAACCCACGAAAGCAAACTCAAAA 1417
|||
QY 539 GAGAAGAAAGCAAAACCAAGAACCTCTCTGATGRCGTTTCTCAGACAAATG 588
|||
Db 1418 GAGAAGAAAGCAAAACCAAGAACCTCTCTGATGRCGTTTCTCAGACAAATG 1467
|||

RESULT 7
US-08-892-695-12
; Sequence 12, Application US/08892695A
; Patent No. 6808878
; GENERAL INFORMATION:
; APPLICANT: Gray, Joe W
; APPLICANT: Collins, Collin
; APPLICANT: Hwang, Soo In
; APPLICANT: Godfrey, Tony
; APPLICANT: Kowel, David
; APPLICANT: Rommens, Johanna
; TITLE OF INVENTION: GENES FROM THE 20Q13 AMPLICON AND THEIR USES
; FILE REFERENCE: 2500.124US3
; CURRENT APPLICATION NUMBER: US/08/892,695A
; CURRENT FILING DATE: 1997-07-15
; EARLIER APPLICATION NUMBER: 08/785,532
; EARLIER FILING DATE: 1997-01-17
; EARLIER APPLICATION NUMBER: 08/731,499
; EARLIER FILING DATE: 1996-10-16
; EARLIER APPLICATION NUMBER: 08/680,395
; EARLIER FILING DATE: 1996-07-15
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 12
; LENGTH: 3066
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:lbl
US-08-892-695-12

Query Match 23.2%; Score 167.6; DB 4; Length 3066;
Best Local Similarity 98.2%; Pred. No. 1.9e-44;
Matches 167; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 419 AGTGTGTGTGTGAGTCACCAGTAGAGATTATAAAGTCCAAGGAAGTAGAATCAGCCTTAC 478
|||
Db 1433 ATGTGTGTGTGAGTCACCAGTAGAGATTATAAAGTCCAAGGAAGTAGAATCAGCCTTAC 1492
|||
QY 479 AAACAGTGGACCTCAACGAAGGAGATGCTGCACCTTGAACCCACGAAAGCAAACTCAAAA 538
|||
Db 1493 AAACAGTGGACCTCAACGAAGGAGATGCTGCACCTTGAACCCACGAAAGCAAACTCAAAA 1552
|||
QY 539 GAGAAGAAAGCAAAACCAAGAACCTCTCTGATGRCGTTTCTCAGACAAATG 588
|||
Db 1553 GAGAAGAAAGCAAAACCAAGAACCTCTCTGATGRCGTTTCTCAGACAAATG 1602
|||
```


RESULT 8

US-09-949-016-4308
; Sequence 4308, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4308
; LENGTH: 3629
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-4308

Query Match 23.2%; Score 167.6; DB 4; Length 3629;
Best Local Similarity 98.2%; Pred. No. 2.1e-44;
Matches 167; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 419 AGTGCTGTGTCAGTCACAGGAGATATATAAGTCCAGGAGTAGAATCAGCCTTAC 478
DB 1298 ATGTGCTGTGTCAGTCACAGGAGATATATAAGTCCAGGAGTAGAATCAGCCTTAC 1357
QY 479 AAACAGTGACCTCAACGAGGAGATGCTGCACCTGAACCCACGAAAGCAACTCAAAA 538
DB 1358 AAACAGTGACCTCAACGAGGAGATGCTGCACCTGAACCCACGAAAGCAACTCAAAA 1417
QY 539 GAGAGAAAGCAAAACCAAGAACCTCTCTGATGCGTTTCTCAGACAAATG 588
DB 1418 GAGAGAAAGCAAAACCAAGAACCTCTCTGATGCGTTTCTCAGACAAATG 1467

RESULT 9

US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION NUMBER: US/07/935,313
; FILING DATE: EP 91 114 300.6
; APPLICATION NUMBER: 26-AUG-1991

ATTORNEY/AGENT INFORMATION:

; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-Fls
US-08-232-463-14

Query Match 5.8%; Score 41.8; DB 1; Length 7218;
Best Local Similarity 1.6%; Pred. No. 0.012;
Matches 5; Conservative 187; Mismatches 125; Indels 0; Gaps 0;

QY 406 GCGAAATAATCCAGGTGCTGTGAGTCACAGTAGAGATTATAAAGTCCAAAGGAAGTA 465
DB 1378 RRR 1319
QY 466 GAATCAGCCTTACAAACAGTGGACCTCAACGAAGAGATGTCACCTGAACCCACWAAA 525
DB 1318 RRR 1259
QY 526 CGGAACTCAAAAGAGAAGAAAGCAACCAAGAGACCTCTCTGATGRCGTTTCTCAGCAA 585
DB 1258 RRR 1199
QY 586 ATGTAAGCCCTTACTTCCAGTATAGGAACTAAGATACCTAGAGCGGCTTTTGGGAA 645
DB 1198 RRR 1139
QY 646 CAATGGGCTCATGCCACAGTAGTAGGAGACATAATTGCTGCTGTATGGAATGTG 705
DB 1138 RRR 1079
QY 706 AATGGAATATGATGTC 722
DB 1078 RRRRRRRRRRRATCGC 1062

RESULT 10

US-09-248-796A-3540
; Sequence 3540, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 3540
; LENGTH: 1728
; TYPE: DNA
; ORGANISM: Candida albicans
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (157), (170), (175)
; OTHER INFORMATION: Identity of nucleotide sequences at the above locations are unknown
US-09-248-796A-3540

```
Query Match          5.0%; Score 36; DB 4; Length 1728;
Best Local Similarity 55.9%; Pred. No. 0.44;
Matches 66; Conservative 1; Mismatches 51; Indels 0; Gaps 0;

QY 439 GTAGAGATTATAAGTCCAGGAGTAGAATCAGCCTTCAACACAGTGACCTCAACGAA 498
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 871 GAAGTGATTGAATTTCTGATGATGTACAAAGCCACAGAAGAACCAAGAAATCAACAA 930

QY 499 GGAGATGCTGCACCTCAACCCAGCAGCGAACTCAAAAGAGAAAGCAACCAAA 556
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 931 GTAATTGAAGAAATTGAAGAAACTGAAGAACTGAAGAAATTTGAAGAA 988

RESULT 11
US-09-949-016-162024/c
; Sequence 162024, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949, 016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 162024
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-162024

Query Match          4.8%; Score 34.8; DB 4; Length 601;
Best Local Similarity 57.3%; Pred. No. 0.59;
Matches 63; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 352 ACTTTGTAACAACATGTTTGGGGGAATGTTTGTTCATTTTCTTATTACCTGCGAAA 411
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 448 AGTGTGTGTACTGTGAATCGTCTATTTTATCAATGAAACATTTATAAAGTGAATA 389

QY 412 ATAATCCAGTGTGTGTGAGTCACCAAGTAGAGATTATAAGTCCAAAGGA 461
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 388 ATAATGCCCTTAATGTGAGGGTTTGTAAATGGTCTTATTAAAGACCAAGA 339

RESULT 12
US-09-949-016-204212
; Sequence 204212, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949, 016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 204212
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-204212

Query Match          5.0%; Score 36; DB 4; Length 1728;
Best Local Similarity 55.9%; Pred. No. 0.44;
Matches 66; Conservative 1; Mismatches 51; Indels 0; Gaps 0;

QY 439 GTAGAGATTATAAGTCCAGGAGTAGAATCAGCCTTCAACACAGTGACCTCAACGAA 498
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 871 GAAGTGATTGAATTTCTGATGATGTACAAAGCCACAGAAGAACCAAGAAATCAACAA 930

QY 499 GGAGATGCTGCACCTCAACCCAGCAGCGAACTCAAAAGAGAAAGCAACCAAA 556
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 931 GTAATTGAAGAAATTGAAGAACTGAAGAACTGAAGAAATTTGAAGAA 988

RESULT 11
US-09-949-016-162024/c
; Sequence 162024, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949, 016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 162024
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-162024

Query Match          4.8%; Score 34.8; DB 4; Length 601;
Best Local Similarity 57.3%; Pred. No. 0.59;
Matches 63; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 352 ACTTTGTAACAACATGTTTGGGGGAATGTTTGTTCATTTTCTTATTACCTGCGAAA 411
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 448 AGTGTGTGTACTGTGAATCGTCTATTTTATCAATGAAACATTTATAAAGTGAATA 389

QY 412 ATAATCCAGTGTGTGTGAGTCACCAAGTAGAGATTATAAGTCCAAAGGA 461
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 388 ATAATGCCCTTAATGTGAGGGTTTGTAAATGGTCTTATTAAAGACCAAGA 339

RESULT 12
US-09-949-016-204212
; Sequence 204212, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949, 016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 204212
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-204212

Query Match          4.8%; Score 34.8; DB 4; Length 601;
Best Local Similarity 50.6%; Pred. No. 0.59;
Matches 81; Conservative 1; Mismatches 78; Indels 0; Gaps 0;

QY 523 GAAGCGAACTCAAAAGAGAAAGCAAAACCAAGAACCTCTCTGATGCGTTTCTCAGA 582
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5  GAGCGAGACTCACTCTCAAAAAAACAACAAAAAACAAGTGTGTGG 64

QY 583 CAAATGTTAAGCCCTTACTTCCAGTATAGAAAACTTAAGATACCTAGAGCGGCTTTGG 642
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 65  AGAAATCTGAACCCCTTGTGCGCTATTGGTGGGAATGTAATAATGGAGTAGCAGCTATGGA 124

QY 643 GAACATGGGCTCATGCGACACAGGTAGTAGGACATAATT 582
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 125 AAACAGTGTGTCATTTCTCAAAATCCTGGACATAGAATT 164

RESULT 13
US-09-949-016-4532
; Sequence 4532, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949, 016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4532
; LENGTH: 9668
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-4532

Query Match          4.8%; Score 34.8; DB 4; Length 9668;
Best Local Similarity 57.3%; Pred. No. 3;
Matches 63; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 352 ACTTTGTAACAACATGTTTGGGGGAATGTTTGTTCATTTTCTTATTACCTGCGAAA 411
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 9335 AGTGTGTGTACTGTGAATCGTCTATTTTATCAATGAAACATTTATAAAGTGAATA 9394

QY 412 ATAATCCAGTGTGTGTGAGTCACCAAGTAGAGATTATAAGTCCAAAGGA 461
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 9395 ATAATGCCCTTAATGTGAGGGTTTGTAAATGGTCTTATTAAAGACCAAGA 9444

RESULT 14
US-09-949-016-17508/c
; Sequence 17508, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949, 016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
```

Search completed: May 5, 2005, 07:03:13
Job time : 170.759 secs

This Page Blank (uspto)

GenCore version 5.1.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 5, 2005, 03:59:12 ; Search time 993.612 Seconds
(without alignments)
4449.254 Million cell updates/sec

Title: US-08-731-499-2

Perfect score: 723

Sequence: 1 TGGAGCTGTCATGCTTACC.....TGAATGGAATGATTGCG 723

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 5654200 seqs, 3057283753 residues

Total number of hits satisfying chosen parameters: 11308400

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications NA.*

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq.*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq.*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq.*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq.*
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq.*
15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq.*
16: /cgn2_6/ptodata/2/pubpna/US10D_PUBCOMB.seq.*
17: /cgn2_6/ptodata/2/pubpna/US10E_PUBCOMB.seq.*
18: /cgn2_6/ptodata/2/pubpna/US10F_PUBCOMB.seq.*
19: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
20: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq.*
21: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
22: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	722.2	99.9	723	8	US-08-731-499-2
2	167.6	23.2	349	13	Sequence 889, App
3	167.6	23.2	2813	17	US-10-040-739-889
4	167.6	23.2	2813	17	US-10-172-118-937
5	167.6	23.2	2813	17	US-10-240-425-385
6	167.6	23.2	2813	17	US-10-342-887-937
7	167.6	23.2	2813	18	US-10-723-860-1812
8	167.6	23.2	3066	8	US-08-731-499-12
9	166	23.0	3418	18	US-10-723-860-6167
10	88.6	12.3	383	9	US-10-491-213-85
11	41.2	5.7	957	9	US-09-960-352-10031
					Sequence 845, App

12	41.2	5.7	957	11	US-09-938-842A-845	Sequence 845, App
c 13	39.4	5.4	103464	18	US-10-322-281-323	Sequence 323, App
c 14	39	5.4	3673778	16	US-10-312-841-1	Sequence 1, Appl
c 15	37.4	5.2	408	18	US-10-275-323A-13	Sequence 13, Appl
c 16	37.2	5.1	386	18	US-10-425-115-134418	Sequence 134418,
c 17	37.2	5.1	17534	17	US-10-257-166-108	Sequence 108, App
c 18	37.2	5.1	83836	13	US-10-087-192-1849	Sequence 1849, Ad
c 19	36.6	5.1	3353	18	US-10-357-930-25269	Sequence 23269, A
c 20	36.6	5.1	3353	18	US-10-357-930-27819	Sequence 27819, A
c 21	36.4	5.0	559	13	US-10-027-632-186364	Sequence 186364,
c 22	36.4	5.0	559	13	US-10-027-632-186365	Sequence 186365,
c 23	36.4	5.0	559	17	US-10-027-632-186364	Sequence 186364,
c 24	36.4	5.0	559	17	US-10-027-632-186365	Sequence 186365,
c 25	36.4	5.0	1113	13	US-10-027-632-260191	Sequence 260191,
c 26	36.4	5.0	1113	13	US-10-027-632-260191	Sequence 260191,
c 27	36.2	5.0	198522	13	US-10-087-192-244	Sequence 244, App
c 28	36	5.0	598	16	US-10-029-386-451	Sequence 451, App
c 29	36	5.0	885	17	US-10-371-264-83	Sequence 83, Appl
c 30	36	5.0	885	17	US-10-371-099-379	Sequence 379, App
c 31	36	5.0	885	17	US-10-371-122-379	Sequence 379, App
c 32	36	5.0	885	17	US-10-373-567-83	Sequence 83, Appl
c 33	36	5.0	885	17	US-10-628-088-379	Sequence 379, App
c 34	36	5.0	885	19	US-10-831-780-379	Sequence 379, App
c 35	36	5.0	13215	17	US-10-371-264-96	Sequence 96, Appl
c 36	36	5.0	13215	17	US-10-371-099-20	Sequence 20, Appl
c 37	36	5.0	13215	17	US-10-373-567-96	Sequence 96, Appl
c 38	36	5.0	13215	17	US-10-373-567-96	Sequence 96, Appl
c 39	36	5.0	13215	17	US-10-628-088-20	Sequence 20, Appl
c 40	36	5.0	13215	19	US-10-831-780-20	Sequence 20, Appl
c 41	36	5.0	13335	18	US-10-789-400-1	Sequence 1, Appli
c 42	36	5.0	14083	18	US-10-789-400-3	Sequence 3, Appli
c 43	35.8	5.0	56339	17	US-10-085-117-280	Sequence 280, App
c 44	35.4	4.9	811	14	US-10-184-644-414	Sequence 414, App
c 45	35.4	4.9	811	14	US-10-184-634-414	Sequence 414, App

ALIGNMENTS

RESULT 1
US-08-731-499-2
; Sequence 2, Application US/08731499
; Publication No. US20030148270A1
; GENERAL INFORMATION:
; APPLICANT: GRAY, Joe W.
; APPLICANT: COLLINS, Colin
; APPLICANT: HWANG, Soo-In
; APPLICANT: GODFREY, Tony
; APPLICANT: KOWBEL, David
; APPLICANT: ROMMENS, Johanna
; TITLE OF INVENTION: GENES FROM THE 20q13 AMPLICON AND THEIR
; TITLE OF INVENTION: USES
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/731.499
; FILING DATE: 16-OCT-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/680,395
; FILING DATE: 15-JUL-1996
; ATTORNEY/AGENT INFORMATION:

```

/ NAME: Hunter, Tom
/ REGISTRATION NUMBER: 38,498
/ REFERENCE/DOCKET NUMBER: 23070-068910
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (415) 576-0200
/ TELEFAX: (415) 576-0300
/ INFORMATION FOR SEQ ID NO: 2:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 723 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: cDNA
/ FEATURE:
/ NAME/KEY: -
/ LOCATION: 1..723
/ OTHER INFORMATION: /note= "cDNA clone 1b11 of 3.5kb
/ OTHER INFORMATION: transcript"
/
US-08-731-499-2

Query Match          99.9%; Score 722.2; DB 8; Length 723;
Best Local Similarity 100.0%; Pred. No. 7.6e-211;
Matches 723; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1  TGAAGCTGTCATGGTTACCGTCTCTAACGTTGGACTCTTAAGAAAATGATATTCTCTGG 60
DB      1  TGAAGCTGTCATGGTTACCGTCTCTAACGTTGGACTCTTAAGAAAATGATATTCTCTGG 60

QY     61  TTTCTAGACAGGCCAATGTAAATTCACCTACGTGGCAGATTAAAGAGGTGGGCTTACTAG 120
DB     61  TTTCTAGACAGGCCAATGTAAATTCACCTACGTGGCAGATTAAAGAGGTGGGCTTACTAG 120

QY    121  ATTTGATTGGGTATTGAGCATGCTCTCAATGACAGTCCCAAAAAGGACCTCTTATCCGT 180
DB    121  ATTTGATTGGGTATTGAGCATGCTCTCAATGACAGTCCCAAAAAGGACCTCTTATCCGT 180

QY    181  TCTTCCCTTTGGGGAGGGCTTTTGGCACTTCATGTCAATGTGGCAGTTGAGCTTGGAA 240
DB    181  TCTTCCCTTTGGGGAGGGCTTTTGGCACTTCCATGTCAATGTGGCAGTTGAGCTTGGAA 240

QY    241  ATTGGTGC GTGTACAACATAAGCATTTACTTCTCAAGATGTGCTGTAGAAATGGTC 300
DB    241  ATTGGTGC GTGTACAACATAAGCATTTACTTCTCAAGATGTGCTGTAGAAATGGTC 300

QY    301  ATAGATTCAAAACTGTAGTACTATGTGGACAGGGGGCGACCAAGGACCCACATTTGTAA 360
DB    301  ATAGATTCAAAACTGTAGTACTATGTGGACAGGGGGCGACCAAGGACCCACATTTGTAA 360

QY    361  AACATGTTTGGGGGAATGTTTTGTTTTTCTATTTTCTATTACCTGGCAAAATAATCCAG 420
DB    361  AACATGTTTGGGGGAATGTTTTGTTTTTCTATTTTCTATTACCTGGCAAAATAATCCAG 420

QY    421  GTGGTGTGTAGTCAACAGTAGAGATTATAAAGTCCAAGGAAGTAGAATCAGCCTTACAA 480
DB    421  GTGGTGTGTAGTCAACAGTAGAGATTATAAAGTCCAAGGAAGTAGAATCAGCCTTACAA 480

QY    481  ACAGTGGAACCTCAACGAGGAGATGCTGCACCTGAACCCACWGAAGCGAAACTCAAGA 540
DB    481  ACAGTGGAACCTCAACGAGGAGATGCTGCACCTGAACCCACWGAAGCGAAACTCAAGA 540

QY    541  GAAGAAAGCAAAACCAAGAACCTCTCTCATGTCGTTTTCTCAGACAAATGGTAAGCCCTTTA 600
DB    541  GAAGAAAGCAAAACCAAGAACCTCTCTCATGTCGTTTTCTCAGACAAATGGTAAGCCCTTTA 600

QY    601  CTTCCAGTATAGGAAACCTAAGATACCTAGACGGGCTTTTGGGAACCAATGGGCTCATGCC 660
DB    601  CTTCCAGTATAGGAAACCTAAGATACCTAGACGGGCTTTTGGGAACCAATGGGCTCATGCC 660

QY    661  ACAGGTAGTAGGACACATAATTGTAGCTGGTGTGATGGAATGTCAATGGAATATGGATT 720
DB    661  ACAGGTAGTAGGACACATAATTGTAGCTGGTGTGATGGAATGTCAATGGAATATGGATT 720

QY    721  GCG 723

```

```

; Publication No. US20030224374A1
; GENERAL INFORMATION:
; APPLICANT: Dai, Hongyue
; APPLICANT: He, Yudong
; APPLICANT: Linsley, Peter
; APPLICANT: Mao, Mao
; APPLICANT: Roberts, Chris
; APPLICANT: Van 't Veer, Laura
; APPLICANT: Van de Vijver, Marc
; APPLICANT: Bernards, Rene
; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
; CURRENT APPLICATION NUMBER: US/10/172,118
; CURRENT FILING DATE: 2002-06-14
; PRIOR APPLICATION NUMBER: 60/380,770
; PRIOR FILING DATE: 2002-05-14
; NUMBER OF SEQ ID NOS: 2699
; SEQ ID NO 937
; LENGTH: 2813
; TYPE: DNA
; ORGANISM: Homo sapiens
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: NM_003657
; DATABASE ENTRY DATE: 2001-06-18
US-10-172-118-937

Query Match      23.2%; Score 167.6; DB 17; Length 2813;
Best Local Similarity 98.2%; Pred. No. 3.4e-40;
Matches 167; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 419 AGGTGGTGTGTGAGTCCACCTGAGATTTATAAAGTCCAAAGGAGTAGAATCAGCCTTAC 478
Db 1298 ATGTGGTGTGTGAGTCCACCTGAGATTTATAAAGTCCAAAGGAGTAGAATCAGCCTTAC 1357

Qy 479 AAACAGTGACCTCAACGAAGGAGTGTGACCTGAAACCCACGAAAGCGAAATCTCAAAA 538
Db 1358 AAACAGTGACCTCAACGAAGGAGTGTGACCTGAAACCCACGAAAGCGAAATCTCAAAA 1417

Qy 539 GAGAAGAAAGCAACCAAGAACCTCTCTGATGCGTTTCTCAGACAAATG 588
Db 1418 GAGAAGAAAGCAACCAAGAACCTCTCTGATGCGTTTCTCAGACAAATG 1467

RESULT 4
US-10-425-385
; Sequence 385, Application US/10240425
; Publication No. US20040033502A1
; GENERAL INFORMATION:
; APPLICANT: Williams, Amanda
; APPLICANT: Boland, Joseph F.
; APPLICANT: Lord, Reginald V.
; APPLICANT: Alvarez, Chris
; APPLICANT: Wetzel, Jon C.
; APPLICANT: Scherf, Uwe
; APPLICANT: Vockley, Joseph G.
; TITLE OF INVENTION: Gene Expression Profiles in Esophageal Tissue
; FILE REFERENCE: 44921-5026
; CURRENT APPLICATION NUMBER: US/10/240,425
; CURRENT FILING DATE: 2002-09-30
; PRIOR APPLICATION NUMBER: PCT/US01/09847
; PRIOR FILING DATE: 2001-03-28
; PRIOR APPLICATION NUMBER: US 60/193,446
; PRIOR FILING DATE: 2000-03-31
; NUMBER OF SEQ ID NOS: 1588
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 385
; LENGTH: 2813
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20040033502A1 AF041260
US-10-240-425-385

```

```

Query Match      23.2%; Score 167.6; DB 17; Length 2813;
Best Local Similarity 98.2%; Pred. No. 3.4e-40;
Matches 167; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 419 AGGTGGTGTGTGAGTCCACCTGAGATTTATAAAGTCCAAAGGAGTAGAATCAGCCTTAC 478
Db 1298 ATGTGGTGTGTGAGTCCACCTGAGATTTATAAAGTCCAAAGGAGTAGAATCAGCCTTAC 1357

Qy 479 AAACAGTGACCTCAACGAAGGAGTGTGACCTGAAACCCACGAAAGCGAAATCTCAAAA 538
Db 1358 AAACAGTGACCTCAACGAAGGAGTGTGACCTGAAACCCACGAAAGCGAAATCTCAAAA 1417

Qy 539 GAGAAGAAAGCAACCAAGAACCTCTCTGATGCGTTTCTCAGACAAATG 588
Db 1418 GAGAAGAAAGCAACCAAGAACCTCTCTGATGCGTTTCTCAGACAAATG 1467

RESULT 5
US-10-342-887-937
; Sequence 937, Application US/10342887
; Publication No. US20040058340A1
; GENERAL INFORMATION:
; APPLICANT: Dai, Hongyue
; APPLICANT: He, Yudong
; APPLICANT: Linsley, Peter S.
; APPLICANT: Mao, Mao
; APPLICANT: Roberts, Christopher J.
; APPLICANT: Van 't Veer, Laura Johanna
; APPLICANT: Van de Vijver, Marc J.
; APPLICANT: Bernards, Rene
; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
; FILE REFERENCE: 9301-188-999
; CURRENT APPLICATION NUMBER: US/10/342,887
; CURRENT FILING DATE: 2003-01-15
; PRIOR APPLICATION NUMBER: 60/298,918
; PRIOR FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: 60/380,710
; PRIOR FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: 10/172,118
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 2699
; SEQ ID NO 937
; LENGTH: 2813
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-342-887-937

Query Match      23.2%; Score 167.6; DB 17; Length 2813;
Best Local Similarity 98.2%; Pred. No. 3.4e-40;
Matches 167; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 419 AGGTGGTGTGTGAGTCCACCTGAGATTTATAAAGTCCAAAGGAGTAGAATCAGCCTTAC 478
Db 1298 ATGTGGTGTGTGAGTCCACCTGAGATTTATAAAGTCCAAAGGAGTAGAATCAGCCTTAC 1357

Qy 479 AAACAGTGACCTCAACGAAGGAGTGTGACCTGAAACCCACGAAAGCGAAATCTCAAAA 538
Db 1358 AAACAGTGACCTCAACGAAGGAGTGTGACCTGAAACCCACGAAAGCGAAATCTCAAAA 1417

Qy 539 GAGAAGAAAGCAACCAAGAACCTCTCTGATGCGTTTCTCAGACAAATG 588
Db 1418 GAGAAGAAAGCAACCAAGAACCTCTCTGATGCGTTTCTCAGACAAATG 1467

RESULT 6
US-10-723-860-1812
; Sequence 1812, Application US/10723860
; Publication No. US20040253606A1
; GENERAL INFORMATION:
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsburg, Wendy M.
; APPLICANT: Zlotnik, Albert
; TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &

```

```

; TITLE OF INVENTION: Methods for Screening for Soft Tissue Sarcoma Modulators
; FILE REFERENCE: 05882.0193.NPUS01
; CURRENT APPLICATION NUMBER: US/10/723,860
; CURRENT FILING DATE: 2003-11-26
; PRIOR APPLICATION NUMBER: 60/429,739
; PRIOR FILING DATE: 2002-11-26
; NUMBER OF SEQ ID NOS: 8393
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1812
; LENGTH: 2813
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-723-860-1812

Query Match      23.2%; Score 167.6; DB 18; Length 2813;
Best Local Similarity 98.2%; Pred. No. 3.4e-40;
Matches 167; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 419 AGTGGTGTGTGAGTCACCAGTAGAGATTATAAAGTCCAAGGAAGTAGAATCAGCCTTAC 478
DB 1298 ATGTGGTGTGTGAGTCACCAGTAGAGATTATAAAGTCCAAGGAAGTAGAATCAGCCTTAC 1357

QY 479 AAACAGTGGACCTCAACGAAGGAGATGCTGCACCTGAACCCACGACGAAAGCGAACTCAAAA 538
DB 1358 AAACAGTGGACCTCAACGAAGGAGATGCTGCACCTGAACCCACGACGAAAGCGAACTCAAAA 1417

QY 539 GAGAAGAAAGCAAAACCAAGAACCTCTCTGATGTCGTTTCTCAGACAAATG 588
DB 1418 GAGAAGAAAGCAAAACCAAGAACCTCTCTGATGTCGTTTCTCAGACAAATG 1467

RESULT 7
US-08-731-499-12
; Sequence 12, Application US/08731499
; Publication No. US20030148270A1
; GENERAL INFORMATION:
; APPLICANT: GRAY, Joe W.
; APPLICANT: COLLINS, Colin
; APPLICANT: HWANG, Soo-In
; APPLICANT: GODFREY, Tony
; APPLICANT: KOMBEL, David
; APPLICANT: ROMMENS, Johanna
; TITLE OF INVENTION: GENES FROM THE 20q13 AMPLICON AND THEIR
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/731,499
; FILING DATE: 16-OCT-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/680,395
; FILING DATE: 15-JUL-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Hunter, Tom
; REGISTRATION NUMBER: 38,498
; REFERENCE/DOCKET NUMBER: 23070-068910
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:

```

```

; LENGTH: 3066 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: -
; LOCATION: 1..3066
; OTHER INFORMATION: /note= "1b1"
US-08-731-499-12

Query Match      23.2%; Score 167.6; DB 8; Length 3066;
Best Local Similarity 98.2%; Pred. No. 3.6e-40;
Matches 167; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 419 AGTGGTGTGTGAGTCACCAGTAGAGATTATAAAGTCCAAGGAAGTAGAATCAGCCTTAC 478
DB 1433 ATGTGGTGTGTGAGTCACCAGTAGAGATTATAAAGTCCAAGGAAGTAGAATCAGCCTTAC 1492

QY 479 AAACAGTGGACCTCAACGAAGGAGATGCTGCACCTGAACCCACGACGAAAGCGAACTCAAAA 538
DB 1493 AAACAGTGGACCTCAACGAAGGAGATGCTGCACCTGAACCCACGACGAAAGCGAACTCAAAA 1552

QY 539 GAGAAGAAAGCAAAACCAAGAACCTCTCTGATGTCGTTTCTCAGACAAATG 588
DB 1553 GAGAAGAAAGCAAAACCAAGAACCTCTCTGATGTCGTTTCTCAGACAAATG 1602

RESULT 8
US-10-723-860-6167
; Sequence 6167, Application US/10723860
; Publication No. US20040253606A1
; GENERAL INFORMATION:
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsburg, Wendy M.
; APPLICANT: Zlotnik, Albert
; TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &
; FILE REFERENCE: 05882.0193.NPUS01
; CURRENT APPLICATION NUMBER: US/10/723,860
; CURRENT FILING DATE: 2003-11-26
; PRIOR APPLICATION NUMBER: 60/429,739
; PRIOR FILING DATE: 2002-11-26
; NUMBER OF SEQ ID NOS: 8393
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 6167
; LENGTH: 3418
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1815)..(1815)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (3303)..(3416)
; OTHER INFORMATION: n is a, c, g, or t
US-10-723-860-6167

Query Match      23.2%; Score 167.6; DB 18; Length 3418;
Best Local Similarity 98.2%; Pred. No. 3.8e-40;
Matches 167; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 419 AGTGGTGTGTGAGTCACCAGTAGAGATTATAAAGTCCAAGGAAGTAGAATCAGCCTTAC 478
DB 1346 ATGTGGTGTGTGAGTCACCAGTAGAGATTATAAAGTCCAAGGAAGTAGAATCAGCCTTAC 1405

QY 479 AAACAGTGGACCTCAACGAAGGAGATGCTGCACCTGAACCCACGACGAAAGCGAACTCAAAA 538
DB 1406 AAACAGTGGACCTCAACGAAGGAGATGCTGCACCTGAACCCACGACGAAAGCGAACTCAAAA 1465

QY 539 GAGAAGAAAGCAAAACCAAGAACCTCTCTGATGTCGTTTCTCAGACAAATG 588

```


DB 1466 GAGAAGAAAGCAACCAAGAACCTCTCTGATGGCGTTTCTCAGACAAATG 1515

RESULT 9

US-10-491-213-85
; Sequence 85, Application US/10491213
; Publication No. US20050048490A1
; GENERAL INFORMATION:
; APPLICANT: INCYTE CORPORATION; AZIMZAI, Yalda;
; APPLICANT: BAUGHN, Maria R.; BECHA, Shanya D.;
; APPLICANT: BOWEN, Mark L.; CHAWLA, Narinder K.;
; APPLICANT: ELLIOTT, Vicki S.; EMERLING, Brooke M.;
; APPLICANT: GANDHI, Aneena R.; GIETZEN, Kimberly J.;
; APPLICANT: GORVAD, Ann E.; GRIFFIN, Jennifer A.;
; APPLICANT: HAFALIA, April J.A.; ISON, Craig H.;
; APPLICANT: KABLE, Amy E.; KALAFUS, Daniel P.;
; APPLICANT: LEHR-MASON, Patricia M.; LU, Dying Aina M.;
; APPLICANT: MARQUIS, Joseph P.; NGUYEN, Dannel B.;
; APPLICANT: RAMKUMAR, Jayalaxmi; RICHARDSON, Thomas W.;
; APPLICANT: KAREHTI, Stephanie K.; SWARNAKAR, Anita;
; APPLICANT: TANG, Y. Tom; TRAN, Uyen K.;
; APPLICANT: WARREN, Bridget A.; XU, Yuming;
; APPLICANT: YAO, Monique G.; YUE, Huibin;
; APPLICANT: YUE, Henry
; TITLE OF INVENTION: PROTEINS ASSOCIATED WITH CELL GROWTH, DIFFERENTIATION, AND DEATH
; FILE REFERENCE: PF-1213 USN
; CURRENT APPLICATION NUMBER: US/10/491,213
; PRIOR FILING DATE: 2004-03-26
; PRIOR APPLICATION NUMBER: PCT/US02/31095
; PRIOR FILING DATE: 2002-09-26
; PRIOR APPLICATION NUMBER: US 60/326,389
; PRIOR FILING DATE: 2001-09-28
; PRIOR APPLICATION NUMBER: US 60/327,380
; PRIOR FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: US 60/328,186
; PRIOR FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: US 60/329,690
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/345,384
; PRIOR FILING DATE: 2001-10-26
; PRIOR APPLICATION NUMBER: US 60/348,165
; PRIOR FILING DATE: 2001-10-26
; PRIOR APPLICATION NUMBER: US 60/350,219
; PRIOR FILING DATE: 2001-11-02
; PRIOR APPLICATION NUMBER: US 60/344,518
; PRIOR FILING DATE: 2001-11-09
; PRIOR APPLICATION NUMBER: US 60/345,143
; PRIOR FILING DATE: 2001-11-09
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: PERL Program
; SEQ ID NO 85
; LENGTH: 2020
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No: 4757551CB1

US-10-491-213-85

Query Match 23.0%; Score 166; DB 19; Length 2020;
Best Local Similarity 97.6%; Pred. No. 9e-40;
Matches 166; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
Qy 419 AGTGGTGTGTCAGTCACAGTAGAGATATTAAGTCCAGGAGTAGAATCAGCCTTAC 478
Db 1312 ATGTGGTGTGTCAGTCACAGTAGAGATATTAAGTCCAGGAGTAGAATCAGCCTTAC 1371
Qy 479 AAACAGTGGCCTCAACGAAGAGAGTGTGACCTTGAACCCACGAGGAGAACTCAAAA 538
Db 1372 AAACAGTGGCCTCAACGAAGAGAGTGTGCTCTGACCCACGAGGAGAACTCAAAA 1431
Qy 539 GAGAAGAAAGCAACCAAGAACCTCTCTGATGGCGTTTCTCAGACAAATG 588

DB 1432 GAGAAGAAAGCAACCAAGAACCTCTCTGATGGCGTTTCTCAGACAAATG 1481

RESULT 10

US-09-960-352-10031
; Sequence 10031, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Mengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 10031
; LENGTH: 383
; TYPE: DNA
; ORGANISM: Bos taurus
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (336)
; OTHER INFORMATION: unsure at all n locations
; OTHER INFORMATION: Clone ID: 43-LIB34-002-Q1-E1-C12
US-09-960-352-10031

Query Match 12.3%; Score 88.6; DB 9; Length 383;
Best Local Similarity 62.9%; Pred. No. 2.3e-16;
Matches 168; Conservative 2; Mismatches 91; Indels 6; Gaps 2;
Qy 328 GGACAGGGGGCAGCAAGGCCACCTTTGTAACATGTTTGGGGAATGTTTGT 387
Db 51 GAAAAGGGGGCCCCAGCTCTCGCTCTGGGCAACTGTTCTGAAAAAGTCAGTTAA 110
Qy 388 TTCATTTTCTTATTACCTGGCAAAATAATCCAGGTGGTGTGTGAGTCCACGATGAGATT 447
Db 111 GAGGATTCAGTCCCCACAGGTGCAGAGGAGAAATGTTGTGTGAGTCCACGATGAGGCT 170
Qy 448 ATAAAGTCCAAAGGAGTAGAATCAGCTTACAAACAGTGGACCT---CAACGAAGGAGAT 504
Db 171 GTAAAGTCTGAAGAAGTAGAGTACGCTTACAAACAGTGGATCTCGGCGAAGAGGAGAG 230
Qy 505 GCTGCCTCTGAACCCACGAGGAGCAAACTCAAAAGAGAGAAAGCAAA--CCAAAGAAC 561
Db 231 CGCACCCGATCCCGAGAGAAAGCAACCCAGGAGAGAAAGCCGAGAGCCCGAGGCC 290
Qy 562 TCTCTGATGRCGTTTCTCAGACAAATG 588
Db 291 TCCCTGATGGCGCTCCTCAGACAGATG 317

RESULT 11

US-09-938-842A-845
; Sequence 845, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; FILE REFERENCE: SAME, AND METHODS OF USE
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16

```

; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 845
; LENGTH: 957
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-845

Query Match          5.7%; Score 41.2; DB 9; Length 957;
Best Local Similarity 47.2%; Pred. No. 0.13;
Matches 118; Conservative 2; Mismatches 130; Indels 0; Gaps 0;

QY 382 TTGTTTTTTCATTTTCTTATTACCTGGCAAAATATCCAGGTGGTGTGTGAGTCACCAAGTA 441
Db 358 TGGATGTTTAGAGTCTTTGGACATGACAAATATGGTACTTGTATGGAGGCTTGCCGAAA 417
QY 442 GAGATTATAAGTCCAGGAAGTAGAATCAGCCCTTACAAACAGTGAACCTCAACGAAGGA 501
Db 418 TGGCGTGTCTTCAGGCTATGACGTTGAATCCAGTGTTCAAATGATGCCATTTTGAAGGC 477
QY 502 GATGCTGCACCTGAACCCACGAGGAACTCAAAAGAGAGAAAGCAACCAAGCAACC 561
Db 478 AGTGCAGCAACTGAGGCTATAGAGAAGATCTATCAAGGACAAACAATTAGCCCAATAACT 537
QY 562 TCTCTGATGRCGTTTCTTCAGACAAATGGTAAGCCCTTACTTCCAGTATAGGAACCTAA 621
Db 538 TTTCAGACGAAGTTCGGCCACATCTTGTATGGCACTTGATCAGGTGAAGGAAACATC 597
QY 622 GATACCTAGA 631
Db 598 GAGGATAAGA 607

```

```

RESULT 12
US-09-938-842A-845
; Sequence 845, Application US/09938942A
; Publication No. US20040009476A9
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; FILE REFERENCE: SAME, AND METHODS OF USE
; FILE REFERENCE: SCRIPT1300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 845
; LENGTH: 957
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-845

```

```

Query Match          5.7%; Score 41.2; DB 11; Length 957;
Best Local Similarity 47.2%; Pred. No. 0.13;
Matches 118; Conservative 2; Mismatches 130; Indels 0; Gaps 0;

QY 382 TTGTTTTTTCATTTTCTTATTACCTGGCAAAATATCCAGGTGGTGTGTGAGTCACCAAGTA 441
Db 358 TGGATGTTTAGAGTCTTTGGACATGACAAATATGGTACTTGTATGGAGGCTTGCCGAAA 417
QY 442 GAGATTATAAGTCCAGGAAGTAGAATCAGCCCTTACAAACAGTGAACCTCAACGAAGGA 501
Db 418 TGGCGTGTCTTCAGGCTATGACGTTGAATCCAGTGTTCAAATGATGCCATTTTGAAGGC 477

```

```

QY 502 GATGCTGCACCTGAACCCACGAGCAAACTCAAAAGAGAGAAAGCAACCAAGCAACC 561
Db 478 AGTGCAGCAACTGAGGCTATAGAGAAGATCTATCAAGGACAAACAATTAGCCCAATAACT 537
QY 562 TCTCTGATGRCGTTTCTTCAGACAAATGGTAAGCCCTTACTTCCAGTATAGGAACCTAA 621
Db 538 TTTCAGACGAAGTTCGGCCACATCTTGTATGGCACTTGATCAGGTGAAGGAAACATC 597
QY 622 GATACCTAGA 631
Db 598 GAGGATAAGA 607

RESULT 13
US-10-322-281-323/c
; Sequence 323, Application US/10322281
; Publication No. US20040126762A1
; GENERAL INFORMATION:
; APPLICANT: David W. Morris
; APPLICANT: Marc S. Malandro
; TITLE OF INVENTION: Novel Compositions and Methods in Cancer
; FILE REFERENCE: 529452001000
; CURRENT APPLICATION NUMBER: US/10/322,281
; CURRENT FILING DATE: 2002-12-17
; NUMBER OF SEQ ID NOS: 866
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 323
; LENGTH: 103464
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..- (103464)
; OTHER INFORMATION: n = A,T,C or G
US-10-322-281-323

```

```

Query Match          5.4%; Score 39.4; DB 18; Length 103464;
Best Local Similarity 52.8%; Pred. No. 5.3;
Matches 85; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

QY 107 GGTGGGCTTACTAGATTGATTTGGTGTATTTGAGCATGCTCTGAATGACACATCCCAAAAG 166
Db 94634 GATGGGCTTCAATGATGATATATTTGGCAAAATTTAGTGAATAGAAACAATTTAGAAAGAG 94575
QY 167 GACCTCTTATCCGTTCTTCCCTTGGGGAAGGCTTTTCCCACTTCCATGTCAATGTGGC 226
Db 94574 AAGTTTATTTAGGTTTATTTATTGAGAAATGTTTTTATAGATACAAAGTAATAGT 94515
QY 227 AGTTGAGCTTGGAAATTTGGTGGCTTTGTACAAACATAGCAATT 267
Db 94514 TTTAAAAATTATGACCTTTGAAGGTAGTAGGACATATAAAAAAT 94474

```

```

RESULT 14
US-10-312-841-1/c
; Sequence 1, Application US/10312841
; Publication No. US20030186277A1
; GENERAL INFORMATION:
; APPLICANT: Epigenomics AG
; TITLE OF INVENTION: Diagnose von bedeutenden genetischen Parametern innerhalb des MHC
; FILE REFERENCE: E01/1208/WO
; CURRENT APPLICATION NUMBER: US/10/312,841
; CURRENT FILING DATE: 2002-12-30
; NUMBER OF SEQ ID NOS: 2
; SEQ ID NO 1
; LENGTH: 3673778
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Chemically treated genomic DNA (Homo sapiens)
; NAME/KEY: unsure
; LOCATION: (3294164)

```

US-10-312-841-1

Query Match	5.4%	Score 39	DB 16	Length 3673778
Best Local Similarity	48.8%	Pred. No. 44		
Matches 102	Conservative 1	Mismatches 106	Indels 0	Gaps 0
Qy	410	AAATAATCCAGGTGGTGTGTGAGTCACCGAGTCAGAGATTATAAAGTCCAAAGAAAGTAGAAT	469	
Db	1519348	AAATACTAATATATATATTTTTTTTAACAATCTTACATATATAATTACACCAAAAATAAA	1519289	
Qy	470	CAGCCTTACAAA CAGTGGACCTCAACGAAGAGATGCTGCACCTGAAACCCACACWGAAGCGA	529	
Db	1519288	ATACCTAAATTTATATTTTACATATTTTAA CAAATAATTTTTAAAAAATAATCCATATAAAAAA	1519229	
Qy	530	AACTCAAAAGAGAAAGCAAA CCAAGAACCCTCTGTGATGRCGTTTCTCAGACCAATGG	589	
Db	1519228	AAACAATAATCAAAAAAAAAAACA CAAAAATTTCCATACCAATCCTTAATACATAAA	1519169	
Qy	590	TAGCCCTTACTTTCCAGTATAGAAACC	618	
Db	1519168	CAATCCCTTTACTATCATATAAAATTCACC	1519140	

RESULT 15

```

US-10-275-323A-13/c
; Sequence 13, Application US/10275323A
; Publication No. US20050079491A1
; GENERAL INFORMATION:
; APPLICANT: DONNE-GOUSSE, CAROLE
; APPLICANT: LAUDET, VINCENT
; APPLICANT: HANNI, CATHERINE
; TITLE OF INVENTION: METHOD OF DETECTING AND IDENTIFYING THE PRESENCE OF BIOLOGICAL
; TITLE OF INVENTION: MATERIALS DERIVED FROM BIRDS, AND OLIGONUCLEOTIDES FOR CARRYING
; FILE REFERENCE: 0508-1052
; CURRENT APPLICATION NUMBER: US/10/275,323A
; CURRENT FILING DATE: 2002-11-08
; PRIOR APPLICATION NUMBER: PCT/FR01/01279
; PRIOR FILING DATE: 2001-04-26
; PRIOR APPLICATION NUMBER: FR 00/05850
; PRIOR FILING DATE: 2000-05-09
; NUMBER OF SEQ ID NOS: 96
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 13
; LENGTH: 408
; TYPE: DNA
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: DNA fragment amplified
; OTHER INFORMATION: from the genome of birds
; FEATURE:
; NAME/KEY: modified base
; LOCATION: (43)..(43)
; OTHER INFORMATION: a, c, g, or t
; FEATURE:
; NAME/KEY: modified base
; LOCATION: (57)..(57)
; OTHER INFORMATION: a, c, g, or t
; FEATURE:
; NAME/KEY: modified base
; LOCATION: (71)..(71)
; OTHER INFORMATION: a, c, g, or t
; FEATURE:
; NAME/KEY: modified base
; LOCATION: (134)..(134)
; OTHER INFORMATION: a, c, g, or t
; FEATURE:
; NAME/KEY: modified base
; LOCATION: (212)..(212)
; OTHER INFORMATION: a, c, g, or t
; FEATURE:
; NAME/KEY: modified base
; LOCATION: (214)..(214)

```

```

; OTHER INFORMATION: a, c, g, or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (217)..(217)
; OTHER INFORMATION: a, c, g, or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (226)..(227)
; OTHER INFORMATION: a, c, g, or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (236)..(236)
; OTHER INFORMATION: a, c, g, or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (245)..(245)
; OTHER INFORMATION: a, c, g, or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (253)..(253)
; OTHER INFORMATION: a, c, g, or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (269)..(269)
; OTHER INFORMATION: a, c, g, or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (324)..(324)
; OTHER INFORMATION: a, c, g, or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (330)..(330)
; OTHER INFORMATION: a, c, g, or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (352)..(352)
; OTHER INFORMATION: a, c, g, or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (355)..(355)
; OTHER INFORMATION: a, c, g, or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (357)..(357)
; OTHER INFORMATION: a, c, g, or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (362)..(362)
; OTHER INFORMATION: a, c, g, or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (364)..(364)
; OTHER INFORMATION: a, c, g, or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (406)..(406)
; OTHER INFORMATION: a, c, g, or t
; DS-10-275-323A-13

```

Query Match 5.2%; Score 37.4; DB 19; Length 408;
Best Local Similarity 9.4%; Pred. No. 1.2;
Matches 36; Conservative 204; Mismatches 140; Indels 1

Qy	92	GTGGCAGATTAAAGAGTGGGCTTACTAGATTGGTATTTGAGCATGCTCTGAATG	151
Db	399	SWAKGGRYRDDDBDDKVKYVTRYRDHYDRBNNVNDKTKNRA NRDBDRDTRWR	340
Qy	152	ACAGTCCCAAAAAGGACCTCTATCGTTCCTCCCTTCGGGAAGGGCTTTGGCACATT	211
Db	339	RKDDWYDYNKDKTNDWEHDBBYTRGTRCTGGWGTTCYCTMTAAATAADKWRDHRHD	280
Qy	212	CCATGCTCAATGTGGCAGTTGAGCTTGGAAATTTGGTGGTGTGTACACAATAGCAATTACTT	271

GenCore version 5.1.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 4, 2005, 23:17:50 ; Search time 3060.29 Seconds
(without alignments)
8992.766 Million cell updates/sec

Title: US-08-731-499-2

Perfect score: 723
Sequence: 1 TGGAGCTGTCATGCTTACC.....TGAATGGAATGATGATTCG 723

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: gb_est1:*

2: gb_est2:*

3: gb_hc:*

4: gb_est3:*

5: gb_est4:*

6: gb_est5:*

7: gb_est6:*

8: gb_gss1:*

9: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
c 1	384.2	53.1	504	5	BQ320066 PM3-CT081
2	169.2	23.4	5738	3	CR749643 Homo sapi
3	166	23.0	280	4	BI041215 IL5-NT027
4	166	23.0	837	4	BG290366 602388204
5	127.6	17.6	645	8	BH086142 RPCI-24-3
c 6	123	17.0	381	2	AW453053 UI-H-BW1
7	122.2	16.9	394	1	AA644590 af74a01.r
c 8	104.2	14.4	416	1	AI060049 UI-R-C1-k
c 9	100	13.8	358	1	AI459809 ap17h05.x
10	98.8	13.7	583	6	CB586288 AWGNNUC.N
11	89.8	12.4	359	2	BE485778
12	88.2	12.2	246	2	BE477506
13	86.6	12.0	668	4	BG964319
14	86.6	12.0	892	5	BU841356 AGENCOURT
15	86.6	12.0	2339	3	AK008957 Mus muscu
c 16	81.4	11.3	906	2	BE305008
17	66.2	9.2	902	2	BF581751
18	51.4	7.1	1000	2	BF577458
19	45.6	6.3	322	2	BF853823 MR2-EN009
20	43.8	6.1	681	3	CB423181
21	42.8	5.9	1341	3	CNS0ADBC
22	42.2	5.8	1208	5	EX898592
c 23	41.6	5.8	879	2	BE740812
24	41.4	5.7	922	4	BG961911
					BG961911 602826590

c 25	41.2	5.7	591	8	AQ934160
26	41.2	5.7	755	5	BQ320066
27	41.2	5.7	1256	3	CNS0ACXK
c 28	40.6	5.6	1111	5	BX442513
c 29	40	5.5	391	7	CO954055 UMC-pnata
c 30	39.8	5.5	282	1	AJ671212
c 31	39.8	5.5	1057	9	CL108034
32	39.6	5.5	1163	2	BF691065
33	39.4	5.4	635	4	BI390256
34	39.4	5.4	734	9	CE248517
35	39.2	5.4	600	7	CO879260
36	39.2	5.4	684	9	BI133486
c 37	39	5.4	615	7	CN034417
38	38.6	5.3	535	8	BH402050
c 39	38.6	5.3	722	7	CN216393
c 40	38.6	5.3	755	8	BH723558
41	38.6	5.3	1302	5	BU505079
c 42	38.4	5.3	658	9	BI172820
c 43	38.4	5.3	722	7	CN059701
c 44	38.2	5.3	202	1	AA109272
45	38.2	5.3	689	9	CE115637

ALIGNMENTS

LOCUS BQ320066 504 bp mRNA linear EST 17-MAY-2002
DEFINITION PM3-CT0817-190501-005-d12 CT0817 Homo sapiens cDNA, mRNA sequence.
ACCESSION BQ320066
VERSION BQ320066.1 GI:20926791
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 504)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V.,
O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202653
10737800
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=PM3&tr2=PM3-CT0817-
190501-005-d12&t3=2001-05-19&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 7
High quality sequence stop: 433.
Location/Qualifiers
1..504
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev_stage="Adult"
/clone_lib="CT0817"
/note="Organ: colon; Vector: puc18; Site_1: Sma1; Site_2:
Sma1; A mini-library was made by cloning products derived

SUMMARIES

Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?cl=IL5&t2=IL5-NT0274-230201-371-cl&t3=2001-02-23&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 279.

FEATURES

Location/Qualifiers
1..280
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev_stage="Adult"
/clone_lib="NT0274"
/note="Organ: nervous tumor; Vector: puc18; Site 1: SmaI; Site 2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

ORIGIN

Query Match 23.0%; Score 166; DB 4; Length 280;
Best Local Similarity 97.6%; Pred. No. 8.8e-38;
Matches 166; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 419 AGTGGTGTGAGTCACAGTAGAGATTATAAGTCCAAGAGTAGAATCAGCCTTAC 478
| | | | |
Db 10 ATGTGGTGTGAGACACAGCAGTAGAGATTATAAGTCCAAGAGTAGAATCAGCCTTAC 69
| | | | |
Qy 479 AAACAGTGGACCTCAACGAAGGAGATGCTGCACCTGAAACCCACGACGAACTCAAAA 538
| | | | |
Db 70 AAACAGTGGACCTCAACGAAGGAGATGCTGCACCTGAAACCCACGACGAACTCAAAA 129
| | | | |
Qy 539 GAGAAGAAAGCAAAACCAAGAACCTCTCTGATGRCGTTTCTCAGACAAATG 588
| | | | |
Db 130 GAGAAGAAAGCAAAACCAAGAACCTCTCTGATGRCGTTTCTCAGACAAATG 179
| | | | |

RESULT 4
BG290366
LOCUS
DEFINITION
602388204F1 NIH_MGC_93 Homo sapiens cDNA clone IMAGE:4517041 5',
mRNA sequence.

ACCESSION
BG290366
VERSION
BG290366.1 GI:13047155
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 837)
NIH-MGC http://mgi.nci.nih.gov/
TITLE
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL
Unpublished (1999)
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-roman@nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10409 row: h column: 02
High quality sequence stop: 718.

FEATURES

Location/Qualifiers
1..837
/organism="Homo sapiens"

/mol_type="mRNA"
/db_xref="taxon:9606"
/clone_lib="IMAGE:4517041"
/tissue_type="transitional cell papilloma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_93"
/note="Organ: bladder; Vector: pCMV-SPORT6; Site 1: NotI; Site 2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 1.7 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH_MGC Library."

ORIGIN

Query Match 23.0%; Score 166; DB 4; Length 837;
Best Local Similarity 97.6%; Pred. No. 1.2e-37;
Matches 166; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 419 AGTGGTGTGAGTCACAGTAGAGATTATAAGTCCAAGAGTAGAATCAGCCTTAC 478
| | | | |
Db 292 ATGTGGTGTGAGTCACAGTAGAGATTATAAGTCCAAGAGTAGAATCAGCCTTAC 351
| | | | |
Qy 479 AAACAGTGGACCTCAACGAAGGAGATGCTGCACCTGAAACCCACGACGAACTCAAAA 538
| | | | |
Db 352 AAACAGTGGACCTCAACGAAGGAGATGCTGCCTCTGAAACCCACGACGAACTCAAAA 411
| | | | |
Qy 539 GAGAAGAAAGCAAAACCAAGAACCTCTCTGATGRCGTTTCTCAGACAAATG 588
| | | | |
Db 412 GAGAAGAAAGCAAAACCAAGAACCTCTCTGATGRCGTTTCTCAGACAAATG 461
| | | | |

RESULT 5

BH086142
LOCUS
DEFINITION
BH086142 645 bp DNA linear GSS 18-JUL-2001
RPCI-24-33809.TJ RPCI-24 Mus musculus genomic clone RPCI-24-33809,
genomic survey sequence.

ACCESSION
BH086142
VERSION
BH086142.1 GI:14905750
KEYWORDS
GSS.

SOURCE
Mus musculus (house mouse)

ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
1 (bases 1 to 645)
Zhao, S., Nierman, W., Malek, J., Shatsman, S., Akinret, B., Levins, M.,
Tsagaye, G., Geer, K., Krol, M., Shvartsbeyn, A., Gebregeorgis, E.,
Russell, D., de Jong, P. and Fraser, C.M.

REFERENCE
Mouse BAC End Sequences from Library RPCI-24

Unpublished (1999)

Other GSSs: RPCI-24-33809.TV

Contact: Shaying Zhao

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPCI-24. For BAC library availability, please contact Pieter de Jong (pdejong@mail.cho.org). Clones may be purchased from BACPAC Resources (http://www.choi.org/bacpac/orderingframe.html). BAC end page: http://www.tigr.org/tldb/bac_ends/mouse/bac_end_intro.html
Plate: 338 row: 0 column: 9
Seq primer: SP6
Class: BAC ends.

FEATURES

Location/Qualifiers
1..645
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clones="RPCI-24-33809"
/sex="Male"
/cell_type="Spleen/Brain"

```
/clone_lib="RPCI-24"
/notes="Vector: pTARBAC1; Site 1: BamH1; Site 2: BamH1;
RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
library was cloned in the pTARBAC1 cloning vector at the
BamH1 sites using MboI partially digested male C57BL/6J
DNA."

ORIGIN
Query Match 17.6%; Score 127.6; DB 8; Length 645;
Best Local Similarity 70.5%; Pred. No. 2.8e-26;
Matches 182; Conservative 2; Mismatches 71; Indels 3; Gaps 1;

QY 371 GGGGGAATGTTTGTGTTTTCATTTCTTATTACCTGGCAAAATAATCCAGGTGGTGTG 430
|||
Db 318 GGGTGTGTTTGTGTTTCTTATTACCTGGCAAAATAATCTAGCGGTGTG 377
|||

QY 431 AGTCACCATAGAGATATAAGTCCAAGAGTAGAATCAGCTTTACAAAAGTGGACC 490
|||
Db 378 AGTCACCATAGAGACCGTAAGCTTGGAGAAAGTAGAATCCAGCTTTACAAAAGTGGATC 437
|||

QY 491 TCAACGAAGGAGATGCTGCACCTGACCCACACGACGAACTCAAAAGAGAAAGCA 550
|||
Db 438 TCAGTGA---AGAGACCGACCTGGAACCCACAGACGTAAGTCAAAAGAGAAAGCAAC 494
|||

QY 551 AACCAAGAACCTCTCTGATGTCCTTCTCAGACAAATGGTAAGCCCTTACTTCCAGTAT 610
|||
Db 495 CCGGGAAGACCTCTGATGGCTTCTCAGACAAATGGTAAGCCACTTCTGATGTCC 554
|||

QY 611 AGAAACCTAAGATACCT 628
|||
Db 555 TGCAGAGGGGACATGCCT 572
|||

RESULT 6
AW453053/c
LOCUS
DEFINITION
UI-H-BW1-ama-e-10-0-UI.61 NCI_CGAP_Sub7 Homo sapiens cDNA clone
IMAGE:3069402 3', mRNA sequence.
ACCESSION
AW453053.1 GI:6993829
VERSION
EST.
KEYWORDS
SOURCE
Homo sapiens (human)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 381)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-r@mail.nih.gov
The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. cDNA Library Preparation: M.B. Soares Lab Clone distribution:
NCI-CGAP clone distribution information can be found through the
I.M.A.G.E. Consortium/LINL at:
www-bio.lnlnl.gov/bbrp/image/image.html
Seq primer: M13 Forward
POLYA=Yes.

FEATURES
Location/Qualifiers
1..381
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3069402"
/lab_host="DH10B (Life Technologies)"
/clone_lib="NCI_CGAP_Sub7"
/notes="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not 1; Site 2: Eco RI; NCI_CGAP_Sub7
is a subtracted library derived from NCI_CGAP_Sub6. The
NCI_CGAP_Sub7 library had 12 million recombinants. A

single-stranded DNA preparation of NCI_CGAP_Sub6 was used
as a tracer in a subtractive hybridization with a driver
comprising: the IMAGE pool (NCI_CGAP_kid3 pool 1 LLAM
3334-3337, 3682-3683, 3798-3803 (IMAGE Clones)
132376-132391, 145608-145677, 150052-150285);
NCI_CGAP_Kid5 pool 1 LLAM 3338-3342, 3722-3725, 3776-3778
(IMAGE Clones) 1323912-1325831, 1471368-1472903,
1492104-1493255); NCI_CGAP_Lu5 pool 1 LLAM 3575-3582,
3851-3854 (IMAGE Clones) 1414920-1417991,
1520904-1522439); NCI_CGAP_GC4 pool 1 LLAM 3164-3167,
3716-3720, 3733-3735 (IMAGE Clones)
1257096-1258631, 1469064-1470983, 1475592-1476743);
NCI_CGAP_P22 pool 1 LLAM 2457-2459, 2758-2759, 3062-3068
(IMAGE Clones) 985608-986759, 1101192-1101959,
1217928-1220615); NCI_CGAP_Co10 pool 1 LLAM 2644-2653,
2871-2872 (IMAGE Clones)
1057416-1061255, 1144584-1145351). (6% of the driver
population), plus a pool of 3,840 arrayed clones from
NCI_CGAP_Sub1 (IMAGE Clones) 2708616-2710535) and
NCI_CGAP_Sub2 (IMAGE Clones) 2710536-2712455) (4% of
the driver population), plus a pool of 11,136 clones from
NCI_CGAP_Sub3 (IMAGE Clones) 2712456-2723591) (10% of
the driver population), plus a pool of 5,472 clones from
NCI_CGAP_Sub4 (IMAGE Clones) 2723592-2729326) (40% of the
driver population), plus a pool of 4032 clones from
NCI_CGAP_Sub6 (IMAGE Clones) 2728969-2733190) (40% of the
driver population). Subtraction was performed as
previously described [Bonaldi, Lennon & Soares (1996):
Normalization and Subtraction: Two Approaches to
Facilitate Gene Discovery. Genome Research 6, 791-806.
TAG_TISSUE=brain
TAG_LIB=NCI_CGAP_Brn50
TAG_SEQ=TTTGG"

ORIGIN
Query Match 17.0%; Score 123; DB 2; Length 381;
Best Local Similarity 94.7%; Pred. No. 5.5e-25;
Matches 126; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 419 AGTGTGTGTGTCAGTCACCATAGAGATTATTAAGTCCAGGAGTAGATCAGCCTTAC 478
|||
Db 133 ATGTGTGTGTGAGTCACCATAGAGATTATTAAGTCCAGGAGTAGATCAGCCTTAC 74
|||

QY 479 AAACAGTGACCTCAACGAAGGAGATGCTGACCTGAACCCAGGAGGAACTCAAAA 538
|||
Db 73 AAACAGTGACCTCAACGAAGGAGATGCTGACCTGAACCCAGGAGGAACTCAAAA 14
|||

QY 539 GAGAGAGAGCAA 551
|||
Db 13 AAAAAAAAAAAAAA 1
|||

RESULT 7
AA644590
AA644590
LOCUS
DEFINITION
af74a01.r1 Soares NhhMPu_S1 Homo sapiens cDNA clone IMAGE:1047720
5', mRNA sequence.
ACCESSION
AA644590
VERSION
AA644590.1 GI:2569808
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 394)
Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,
Krieman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M.,
Martin, J., Moore, B., Schellberg, K., Steptoe, M., Tan, F.,
Theising, B., White, Y., Wyllie, T., Waterston, R. and Wilson, R.
WashU-NCI human EST Project
Unpublished (1997)
Contact: Wilson RK
Washington University School of Medicine
```



```
RESULT 9
AI459809/c
LOCUS
DEFINITION
  ap17h05.x1 Schiller oligodendrogloma Homo sapiens cDNA clone
  IMAGE:1955673 3', mRNA sequence.
ACCESSION
  AI459809
VERSION
  AI459809.1 GI:4312690
KEYWORDS
  EST.
SOURCE
  Homo sapiens (human)
ORGANISM
  Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
  1 (bases 1 to 358)
AUTHORS
  Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
  Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M.,
  Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F.,
  Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.
  WashU-NCI Human EST Project
  Unpublished (1997)
  Contact: Wilson RK
  Washington University School of Medicine
  4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
  Tel: 314 286 1800
  Fax: 314 286 1810
  Email: est@watson.wustl.edu
  This clone is available royalty-free through LML; contact the
  IMAGE Consortium (info@image.llnl.gov) for further information.
  Seq primer: -40UP from Gibco
  High quality sequence stop: 267.
FEATURES
  source
    1..358
    /organism="Homo sapiens"
    /mol_type="mRNA"
    /db_xref="taxon:9606"
    /clone="IMAGE:1955673"
    /sex="male"
    /tissue_type="oligodendrogloma"
    /dev_stage="44 years"
    /lab_host="SOLR"
    /clone_lib="Schiller oligodendrogloma"
    /notes="Organ: brain; Vector: pBluescript SK- (Stratagene);
    Site 1: EcoRI; Site 2: XhoI; Double-stranded cDNA was
    prepared from human oligodendrogloma using primer
    5'-GAGAGAGAGAGAGAGAGAACTAGTCTGAGT(18)-3'. An EcoRI
    adaptor was used on the 5' end of the cDNA as follows:
    5'-AATTCGACGAG-3'. The library was size-selected and
    went through one round of amplification. Average insert
    size is 1.7 kb, with a range from 0.4-12 kb. Tumor
    identification by consensus pathology; contains
    chromosome 1p and 19q deletion as determined by CGH. This
    library was constructed by Dr. Martin Schiller (Johns
    Hopkins University)."
```

```
Query Match      13.8%; Score 100; DB 1; Length 358;
Best Local Similarity 89.8%; Pred. No. 3.6e-18;
Matches 106; Conservative 1; Mismatches 11; Indels 0; Gaps 0;

QY 419 AGTGTGTGTGAGTCACCACTAGAGATTATAAGTCCAAGGAGTAGAATCAGCCTTAC 478
      |||||
Db 118 AGGTGGTGTGAGTCACCACTAGAGATTATAAGTCCAAGGAGTAGAATCAGCCTTAC 59

QY 479 AAACAGTGGACCTCAACGAGGAGTGCTGACCTGACCACCCAGCAGGAGCAAACTCAA 536
      |||||
Db 58 AAACAGTGGCCCTCAACGAGGAGTGCTGCTCCCGAACCCCAAAAAA 1

RESULT 10
CB586288
LOCUS
DEFINITION
  CB586288
  AMGNNUC:NRHYS-00221-H3-A W Rat hypothalamus (10471) Rattus
  norvegicus cDNA clone nrhys-00221-h3 5', mRNA sequence.
FEATURES
  source
    1..583
    /organism="Rattus norvegicus"
    /mol_type="mRNA"
    /db_xref="taxon:10116"
    /clone_lib="W Rat hypothalamus (10471)"
    /note="Vector: pSPORT1; Site 1: SalI; Site 2: NotI; W Rat
    hypothalamus adult female Wistar rat avg. insert size 2.3
    kb fraction 6 and 7"
```

```
Query Match      13.7%; Score 98.8; DB 6; Length 583;
Best Local Similarity 72.9%; Pred. No. 9.4e-18;
Matches 124; Conservative 2; Mismatches 44; Indels 0; Gaps 0;

QY 419 AGTGTGTGTGAGTCACCACTAGAGATTATAAGTCCAAGGAGTAGAATCAGCCTTAC 478
      |||||
Db 39 ACGCGGTGTGTGAGCCACCACTAGAGACTGTAAAGGCTTGAGGAAGTAGAATCCACCTTAC 98

QY 479 AAACAGTGGACCTCAACGAGGAGTGCTGACCTGACCACCCAGCAGGAGCAAACTCAA 538
      |||||
Db 99 AAACCGTGGATCTCACTGAGAAAGAACCCAGACTGAGCCACAGATGTGAAAGTCAAG 158

QY 539 GAGAGAAAGCAAAACCAAGAACCTCTCTGATGCGTTTCTCAGACAAATG 588
      |||||
Db 159 AAGAAAGCAAAACCCGAGGAGCCCTCTCTGATGCGTTTCTCAGACAAATG 208

RESULT 11
BE485778
LOCUS
DEFINITION
  BE485778
  BE485778.1 GI:9605311
  EST.
  Bos taurus (cow)
  Bos taurus
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
  Bovinae; Bos.
  1 (bases 1 to 359)
  Sonstegard,T., Capuco,A.V., White,J., Van Tassel,C.P.,
  Connor,E.E., Cho,J., Sultana,R., Shade,L., Wray,J.E., Wells,K.D.
  and Quackenbush,J.
  Analysis of bovine mammary gland EST and functional annotation of
  the Bos taurus gene index
  Mamm. Genome 13 (7), 373-379 (2002)
  22135956
  12140684
  Contact: Sonstegard TS
  USDA, ARS, Beltsville Agricultural Research Center
  Edlg, 200 Rm 2A, Beltsville, MD 20705, USA
  Tel: 301 504 8416
  Fax: 301 504 8414
  Email: tads@lpsl.barc.usda.gov
```

Single pass sequencing. Bases called and alt trimmed with phred v0.980904.e. Vector identified by cross_match with the -minscore 18 and -minmatch 12 options.

PCR Primers

FORWARD: AGGAAACAGCTATGACCAT

BACKWARD: GTTTCCAGCTCAGCAG

Plate: 132 row: N column: 19

Seq primer: ATTAGGTGACACTATAG.

Location/Qualifiers

1. .359

/organism="Bos taurus"

/mol_type="mRNA"

/db_xref="taxon:9913"

/tissue_type="pooled"

/lab_host="DH10B"

/clone_lib="BARC 5BOV"

/note="Vector: pCMV SPORT6; Site 1: NotI; Site 2: SalI; Library made from pooled mRNA isolated from mammary tissues at eight physiological, developmental, and disease states."

ORIGIN

Query Match 12.4%; Score 89.8; DB 2; Length 359;
Best Local Similarity 76.3%; Pred. No. 3.9e-15;
Matches 135; Conservative 2; Mismatches 34; Indels 6; Gaps 2;

QY 418 CAGGTGGTGTGAGTCCACGATAGAGATTATAAGTCCAAAGGAAGTAGAATCAGCCCTTA 477

DB 52 CAGGTGGTGTGAGTCCACGATAGAGCTGTAAAGTCTGAAGAGTAGAGTACGCCCTTA 111

QY 478 CAAACAGTGGACCT---CAACGAAGAGATGCTGCACCTGAACCCACGACGCGAAATC 534

DB 112 CAAACAGTGGATCTCGGCGAAGAGGAGGAGCCACACCCGATCCGCGAAGAGACGCC 171

QY 535 AAACAGAGAGAAAGCAAA---CCAGAACCTCTCTGATGRCGTTTCTCAGACAAATG 588

DB 172 AGGAGAGAAGAACGACGAGGCCCGAGGCCCTCCCTGATGGCGCTCCTCAGACAGATG 228

RESULT 12

BE477506 246 bp mRNA linear EST 27-MAR-2003
LOCUS 161238 BARC 5BOV Bos taurus cDNA 5', mRNA sequence.
DEFINITION BE477506
ACCESSION BE477506.1 GI:9597039
VERSION
KEYWORDS
SOURCE Bos taurus (cow)
ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.
1 (bases 1 to 246)
Sonstegard T., Capuco A.V., White J., Van Tassel C.P., Connor E.E., Cho J., Sultana R., Shade L., Wray J.E., Wells K.D. and Quackenbush J.
Analysis of bovine mammary gland EST and functional annotation of the Bos taurus gene index
Mamm. Genome 13 (7), 373-379 (2002)

22135956
12140684
Contact: Sonstegard TS
USDA, ARS, Beltsville Agricultural Research Center
Bldg. 200 Rm 2A, Beltsville, MD 20705, USA
Tel: 301 504 8416
Fax: 301 504 8414
Email: tad@lpei.barc.usda.gov

Single pass sequencing. Bases called and alt trimmed with phred v0.980904.e. Vector identified by cross_match with the -minscore 18 and -minmatch 12 options.

PCR Primers

FORWARD: AGGAAACAGCTATGACCAT

BACKWARD: GTTTCCAGCTCAGCAG

Plate: 3 row: B column: 3

FEATURES

source

Seq primer: ATTAGGTGACACTATAG.

Location/Qualifiers

1. .246

/organism="Bos taurus"

/mol_type="mRNA"

/db_xref="taxon:9913"

/tissue_type="pooled"

/lab_host="DH10B"

/clone_lib="BARC 5BOV"

/note="Vector: pCMV SPORT6; Site 1: NotI; Site 2: SalI; Library made from pooled mRNA isolated from mammary tissues at eight physiological, developmental, and disease states."

ORIGIN

Query Match 12.2%; Score 88.2; DB 2; Length 246;
Best Local Similarity 75.7%; Pred. No. 1e-14;
Matches 134; Conservative 2; Mismatches 35; Indels 6; Gaps 2;

QY 418 CAGGTGGTGTGAGTCCACGATAGAGATTATAAGTCCAAAGGAAGTAGAATCAGCCCTTA 477

DB 37 CAGGTGGTGTGAGTCCACGATAGAGCTGTAAAGTCTGAAGAGTAGAGTACGCCCTTA 96

QY 478 CAAACAGTGGACCT---CAACGAAGGAGATGCTGCACCTGAACCCACGACGCGAAATC 534

DB 97 CAAACAGTGGATCTCGGCGAAGAGGAGGAGCCACCCGATCCGCGAAGAGACGCC 156

QY 535 AAACAGAGAAAGCAAA---CCAGAACCTCTCTGATGRCGTTTCTCAGACAAATG 588

DB 157 GGGAGAGAGAAACGACGAGGCCCGAGGCCCTCCCTGATGGCGCTCCTCAGACAGATG 213

RESULT 13

BE964319 668 bp mRNA linear EST 12-JUN-2001
LOCUS 602831931F1 NCI_CGAP_Co24 Mus musculus cDNA clone IMAGE:4986798 5', mRNA sequence.
DEFINITION BE964319
ACCESSION BE964319.1 GI:14351956
VERSION
KEYWORDS
SOURCE Mus musculus (house mouse)
ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 668)

NIH-MGC <http://mgc.nci.nih.gov/>.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

Tissue Procurement: Jeffrey E. Green, M.D.

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: L1AM10996 row: m column: 07

High quality sequence stop: 668.

Location/Qualifiers

1. .668

/organism="Mus musculus"

/mol_type="mRNA"

/strain="FVB/N"

/db_xref="taxon:10090"

/clone="IMAGE:4986798"

/lab_host="DH10B (T1 phage-resistant)"

/clone_lib="NCI CGAP Co24"

/note="Organ: colon; Vector: pCMV-SPORT6; Site 1: NotI; Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.

Average insert size 1.6 kb. Constructed by Life Technologies. Note: this is a NCI_CGAP Library."

FEATURES

source

ORIGIN

```
Query Match      12.0%; Score 86.6; DB 4; Length 668;
Best Local Similarity 60.7%; Pred. No. 4.1e-14;
Matches 156; Conservative 2; Mismatches 96; Indels 3; Gaps 1;

QY 332 AGGGGGGAGCAAGAGCCACCTTTGTAAACAATGTTTGGGGGAATGTTTGTTCATCA 391
    |||
Db 207 AGGCAGGGGCCACCTCACTACCGCTGGGAAAGTGTGTTTGGAAAGTCAGTTAAGGAGG 266
    |||
QY 392 TTTTCTTATTAACCTGCAAAATAATCCAGGTGTGTGAGTCACCAAGTAGAGATTATAA 451
    |||
Db 267 ATACACTTTCCACAGGTGCAGAGGAGAACCGGTGTGTGAGTCACCAAGTAGAGACCGTAA 326
    |||
QY 452 AGTCCAAAGGAAGTAGAATCAGCTTACAAACAGTCGACCTCAACGAGGAGATGTCGCAC 511
    |||
Db 327 GGTTCAGGAAGTAGAATCAGCTTACAAACAGTCGATCTCAGTGA--AGAGACCCAGC 383
    |||
QY 512 CTGAACCCACWGAAGCGAAACTCAAAAGAGAGAAAGCAAAACCAAGAACTCTCTGATGR 571
    |||
Db 384 CTGAACCCACAGACGTAAAAGTCAAAAGAGAAAGCAAAACCCCGGAAGACCCCTCTCATGG 443
    |||
QY 572 CGTTTCTCAGACAAATG 588
    |||
Db 444 CGTTTCTCAGACAAATG 460
    |||

RESULT 14
BU8411356
LOCUS
DEFINITION
AGENCOURT_10421743 NCI_CGAP_Co24 Mus musculus cDNA clone
IMAGE:6529081 5', mRNA sequence.

ACCESSION
BU8411356
VERSION
BU8411356.1 GI:24025792
KEYWORDS
Mus musculus (house mouse)
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-x@mail.nih.gov
Tissue Procurement: The Cepko Laboratory
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM14128 row: k column: 01
High quality sequence stop: 642.
Location/Qualifiers
1..892
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="FVB/N"
/clone="IMAGE:6529081"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NCI_CGAP Co24"
/notes="Organ: Colon; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dr.
Average insert size 1.6 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP Library."

FEATURES
source

ORIGIN

Query Match      12.0%; Score 86.6; DB 5; Length 892;
Best Local Similarity 60.7%; Pred. No. 4.4e-14;
Matches 156; Conservative 2; Mismatches 96; Indels 3; Gaps 1;

QY 332 AGGGGGGAGCAAGAGCCACCTTTGTAAACAATGTTTGGGGGAATGTTTGTTCATCA 391
    |||
```

```
Db 268 AGGCAGGGGCCACCTCACTACCGCTGGGAAAGTGTGTTTGGAAAGTCAGTTAAGGAGG 327
    |||
QY 392 TTTTCTTATTAACCTGCAAAATAATCCAGGTGTGTGAGTCACCAAGTAGAGATTATAA 451
    |||
Db 328 ATACACTTTCCACAGGTGCAGAGGAGAACCGGTGTGTGAGTCACCAAGTAGAGACCGTAA 387
    |||
QY 452 AGTCCAAAGGAAGTAGAATCAGCTTACAAACAGTCGACCTCAACGAGGAGATGTCGCAC 511
    |||
Db 388 GGTTCAGGAAGTAGAATCAGCTTACAAACAGTCGATCTCAGTGA--AGAGACCCAGC 444
    |||
QY 512 CTGAACCCACWGAAGCGAAACTCAAAAGAGAGAAAGCAAAACCAAGAACTCTCTGATGR 571
    |||
Db 445 CTGAACCCACAGACGTAAAAGTCAAAAGAGAAAGCAAAACCCCGGAAGACCCCTCTCATGG 504
    |||
QY 572 CGTTTCTCAGACAAATG 588
    |||
Db 505 CGTTTCTCAGACAAATG 521
    |||

RESULT 15
AK008957
LOCUS
DEFINITION
Mus musculus adult male stomach cDNA, RIKEN full-length enriched
library, clone:2210416M21 product:weakly similar to BREAST
CARCINOMA AMPLIFIED SEQUENCE 1 (NOVEL AMPLIFIED IN BREAST CANCER 1)
(AMPLIFIED AND OVEREXPRESSED IN BREAST CANCER) [Homo sapiens], full
insert sequence.
ACCESSION
AK008957
VERSION
AK008957.1 GI:12843457
KEYWORDS
HTC; CAP trapper.
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1
Carninci, P. and Hayashizaki, Y.
AUTHORS
TITLE
High-efficiency full-length cDNA cloning
JOURNAL
Meth. Enzymol. 303, 19-44 (1999)
MEDLINE
99279253
PUBMED
10349636
2
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
20499374
PUBMED
11042159
3
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
Konno, H., Akiyama, J., Nishi, K., Katsunai, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
20530913
PUBMED
11076861
4
The RIKEN Genome Exploration Research Group Phase II Team and the
FANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)
5
The PANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
```

REFERENCE 6 (bases 1 to 2339)
ADACHI, J., AIZAWA, K., AKAHIRA, S., AKIMURA, T., ARAI, A., AONO, H.,
ARAKAWA, T., BONO, H., CARNINCI, P., FUKUDA, S., FUKUNISHI, Y.,
FURUKO, M., HANAGAKI, T., HARA, A., HAYASHI, N., HIRAMOTO, K.,
HIRAKO, T., HORI, F., IMOTANI, K., ISHII, Y., ITOH, M., IZAWA, M.,
KASUKAWA, T., KATO, H., KAWAI, J., KOJIMA, Y., KONNO, H., KOUNDA, M.,
KOYA, S., KURIHARA, C., MATSUYAMA, T., MIYAZAKI, A., NISHI, K.,
NOMURA, K., NUMAZAKI, R., OHNO, M., OKAZAKI, Y., OKIDO, T., OWA, C.,
SATO, H., SATO, R., SAKAI, C., SAKAI, K., SANO, H., SASAKI, D.,
SHIBATA, K., SHIBATA, Y., SHINAGAWA, A., SHIRAKI, T., SOGABE, Y.,
SUZUKI, H., TAGAMI, M., TAGAWA, A., TAKAHASHI, F., TANAKA, T.,
TEJIMA, Y., TOYA, T., YAMAMURA, T., YASUNISHI, A., YOSHIDA, K.,
YOSHINO, M., YURAMATSU, M. and HAYASHIZAKI, Y.
Direct Submission
Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.jp,
URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222,
Fax: 81-45-503-9216)
Please visit our web site (http://genome.gsc.riken.jp/) for further
details.
CDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues. First strand cDNA was primed with a primer
[5', GAGAGAGAGAGATCCAGAGCTCTTTTCTTTTNN 3'], cDNA was
prepared by using trehalose thermo-activated reverse transcriptase
and subsequently enriched for full-length by cap-trapper. Second
strand cDNA was prepared with the primer adapter of sequence [5'
GAGAGAGATTCCTCGATTAAATTAATCCCCCCCC 3']. cDNA was cleaved
with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI.
Host: SOLR.
Location/Qualifiers
1..2339
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM DB:2210416M21"
/db_xref="taxon:10090"
/clone="2210416M21"
/sex="male"
/tissue_type="stomach"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
41..1413
/note="unnamed protein product; putative
weakly similar to BREAST CARCINOMA AMPLIFIED SEQUENCE 1
(NOVEL AMPLIFIED IN BREAST CANCER 1) (AMPLIFIED AND
OVEREXPRESSED IN BREAST CANCER) [Homo sapiens]
(SWISSPROT|Q75363, evidence: FASTY, 56%ID, 70.3%length,
match=1371)"
/codon_start=1
/protein_id="BAB25989.1"
/db_xref="GI:12843458"
/translation="EQTLPATPPESQAEAPADKDFPLNRRFKLDKGRSA
PNSQPKAESGDEQATEAPVGNPHGVSGEDI VDSRGQDVDTLSVSPGDP
EVPYTKEDPQVVDITENSSIMSFFKLVSPNKTETKDPEDYKATKADSVCDGHA
GOMSETAKSKKKLSDSPRLGSLPRKLRHDKTENSPTISANLKSIDKANFTPOETRG
KTKATKSCSPPPPTSEGRDSEKAGPTSLFLGLFWKSKVKEDTLSTGAEANA
VCESPVETVRLVESSLIQTVDLSEETQEPDIDVKVESEKPRKTPLMFLRQMSVRS
SEGI PRSESNVDSKSCOTSNSVEKTPSPPEPAGTAQKNKETSSKDKKSVDKKSA
TENSQKNGKQVRPAPCVQPTVEANAMQTGDKTPKSEKRQSLGFLKGLGPKR
MSDAQVQIDPVSIGPVGKSK"
2317..2322
/note="putative"
2339
/note="putative"

polyA_signal
polyA_site

ORIGIN

Query Match 12.0%; Score 86.6; DB 3; Length 2339;

Best Local Similarity 60.7%; Pred. No. 5.8e-14;
Matches 156; Conservative 2; Mismatches 96; Indels 3; Gaps 1;
QY 332 AGGGGGGCGAGCAAGGACCCCACTTTGTAAACATGTTTGGGGGAATGTTTGTTCATCA 391
DB 737 AGGCAGGGCCCACTACTACCGCTGGGAAAGTTGTTTGGGAAGAAGTCAGTTAAGGAGG 796
QY 392 TTTTCTTATTACCTGGGCAAAATAATCCAGGTGGTGTGTGAGTCACCCAGTAGAGATTATAA 451
DB 797 ATACACTTTCCACAGGTGCAGAGGAGAACGCGGTGTGTGAGTCACCCAGTAGAGACCGTAA 856
QY 452 AGTCCAAGGAAGTAGAATCAAGCTTACAAACAGTGGACCTCAACGAAGGAGATGCTGCAC 511
DB 857 GCCTTGAGGAAGTAGAATCCAGCTTACAAACTGTGGATCTCAGTGA---AGAGACCCAGC 913
QY 512 CTGAACCCCAWGAAGCGAAACTCAAAAGAGAGAAAGCAAAACCAAGAACCTCTCTGTATGR 571
DB 914 CTGAACCCCAWGAAGCGTAAAGTCAAAAGAGAGAAAGCAAAACCCCGGAAGACCCCTCTGTATGG 973
QY 572 CGTTTCTCAGACAAATG 588
DB 974 CGTTTCTCAGACAAATG 990

Search completed: May 5, 2005, 06:53:53
Job time : 3074.62 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 4, 2005, 20:11:25 ; Search time 6286.44 Seconds
(without alignments)
9927.764 Million cell updates/sec

Title: US-08-731-499-5
Perfect score: 1288
Sequence: 1 GAGGCACGAGAGAGAGAA.....AAAAAAAAAAAAAAAAATTC 1288

Scoring table: IDENTITY NUC

Gapop 10_0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*
2: gb_htg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vi.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1288	100.0	1288	6	AR070328
2	1288	100.0	1288	6	BD085729
3	1283.4	99.6	1830	9	BM648148 Homo sapi
4	1265	98.2	3322	6	AX746848 Sequence
5	1265	98.2	3322	9	AK091206 Homo sapi
6	1137	88.3	54161	9	BM276189 Human DNA
7	1137	88.3	172256	2	AC051637 Homo sapi
8	1125	87.3	57493	9	AC004505 Homo sapi
9	1013	78.6	3168	6	C0740312 Sequence
10	909.2	70.6	4502	10	BC079550 Mus muscu
11	909.2	70.6	218037	10	AL731822 Mouse DNA
12	908.2	70.5	4041	10	BC079877 Mus muscu
13	897.2	69.7	3294	10	AF207880 Mus muscu
14	656.4	51.0	2012	6	CQ078477 Sequence
15	568.6	44.1	671	11	BM14871 S212P6046
16	527.4	40.9	1152	10	BC028776 Mus muscu
17	521	40.5	1964	6	CQ069297 Sequence
18	395.2	30.7	3213	10	AF191309 Mus muscu
19	393.6	30.6	2900	10	BC058264 Mus muscu

20	393.6	30.6	2914	10	BC038610	BC038610	Mus muscu
21	393.6	30.6	3215	10	BC017636	BC017636	Mus muscu
22	393.6	30.6	4096	10	BC040215	BC040215	Mus muscu
23	393.6	30.6	193064	10	AC124012	AC124012	Mus muscu
C	24	393.6	30.6	252617	2	AC124231	Mus muscu
C	25	390.4	30.3	183251	2	AC013555	Homo sapi
26	390.4	30.3	189167	9	AC116003	Homo sapi	
27	390.4	30.3	201027	9	AC025105	Homo sapi	
28	389.4	30.2	2933	6	CQ715160	Sequence	
29	388.8	30.2	4459	9	HS0804634		
30	387.2	30.1	4854	6	CQ849970	Sequence	
31	387.2	30.1	4854	9	AK127042	Homo sapi	
32	387.2	30.1	230105	2	AC108227	Rattus no	
C	33	387.2	30.1	239857	2	AC131132	Rattus no
34	387.2	30.1	256029	2	AC094999	Rattus no	
C	35	387.2	30.1	312143	2	AC123509	Rattus no
36	377.8	29.3	2918	6	AR279649	Sequence	
37	377.8	29.3	2918	6	BD079808	Cancer-as	
38	377.8	29.3	2918	9	AF039698	Homo sapi	
39	283.8	22.0	3755	5	AF242292	Danio rer	
40	283.8	22.0	145503	5	BM470238	Zebrafish	
C	41	283.8	22.0	188687	2	BM640405	Danio rer
42	278.4	21.6	230796	2	AC098154	Rattus no	
C	43	276.8	21.5	110000	2	AC108426 ¹	Continuation (2 of
44	275.2	21.4	139125	10	AC104100	Mus muscu	
45	275.2	21.4	179415	2	AC151001	Mus muscu	

ALIGNMENTS

RESULT 1
LOCUS AR070328 1288 bp DNA linear PAT 18-FEB-2000
DEFINITION Sequence 5 from patent US 5892010.
ACCESSION AR070328
VERSION AR070328.1 GI:7221216
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1288)
AUTHORS Gray, J., Collins, C., Hwang, S.-i., Godfrey, T., Kowbel, D. and Rommens, J.
TITLE Genes from the 20Q13 amplicon and their uses
JOURNAL Patent: US 5892010-A 5 06-APR-1999;
FEATURES
Location/Qualifiers
source 1..1288
/organism="unknown"
/mol_type="unassigned DNA"

ORIGIN	Query Match	100.0%;	Score 1288;	DB 6;	Length 1288;
	Best Local Similarity	100.0%;	Pred. No. 0;		
	Matches 1288;	Conservative	0;	Mismatches	0;
				Indels	0;
				Gaps	0;
QY	1	GAGGCACGAGAGAGAGAGAAACCCAGCCCTGGAGCCACACATCTGCTGTGAGCAATGGG	60		
DB	1	GAGGCACGAGAGAGAGAGAAACCCAGCCCTGGAGCCACACATCTGCTGTGAGCAATGGG	60		
QY	61	TGCGCCCTCGCCAAACACAGCCCGCCCTGCATCATCAACCCACTCAGCGCCCTGCAG	120		
DB	61	TGCGCCCTCGCCAAACACAGCCCGCCCTGCATCATCAACCCACTCAGCGCCCTGCAG	120		
QY	121	TCCGTCCTGAACATCACTTGGGCAAAACGACGAGCCCTTGGCGCTCACTTCTGCTCC	180		
DB	121	TCCGTCCTGAACATCACTTGGGCAAAACGACGAGCCCTTGGCGCTCACTTCTGCTCC	180		
QY	181	AGCCCAAGTTCAGCACAATTTCCATGTTCCACAGTCAATCTCAATGTATGACACAG	240		
DB	181	AGCCCAAGTTCAGCACAATTTCCATGTTCCACAGTCAATCTCAATGTATGACACAG	240		
QY	241	CGGCTTTGAGTCTGCTGCCTCCACAGGTGACGCGGTGTCCAGGCGGTACCTGTTTGAG	300		

|||||
241 CCGGTCTTGAGTCTGCTCCACAAGGTGAGCCAGCGGTGTCAGGCGCTACCTGTTGAG 300
|||||
301 AACAGCGATCAGCCCATTTGACCTGACCAAGTCCAAAGCAAGAAAGCCGAGTCTTCGCAA 360
|||||
301 AACAGGATCAGCCCATTTGACCTGACCAAGTCCAAAGCAAGAAAGCCGAGTCTTCGCAA 360
|||||
361 GCACAAATCTTGATGTCTCCACCTCAGAAAGCAGCTCTGTCTGACATCGCGACATGGTC 420
|||||
361 GCACAAATCTTGATGTCTCCACCTCAGAAAGCAGCTCTGTCTGACATCGCGACATGGTC 420
|||||
421 AAAGTCTCTCCAAAGCCACACCCCAAGCCAGCTCTCTCTCCAGGTCCTCCCTCCATG 480
|||||
421 AAAGTCTCTCCAAAGCCACACCCCAAGCCAGCTCTCTCTCCAGGTCCTCCCTCCATG 480
|||||
481 AAGCTGGAATGATGTCTCAGGCGCTTTGAGGATGTCTCCAGTGAAGTCTCAACTTTGCAT 540
|||||
481 AAGCTGGAATGATGTCTCAGGCGCTTTGAGGATGTCTCCAGTGAAGTCTCAACTTTGCAT 540
|||||
541 AAAAGAAAAGCCGCGAGTCCAACTGGAATCTCTCAGCATCTTCTGATTTCTACAAGCCCA 600
|||||
541 AAAAGAAAAGCCGCGAGTCCAACTGGAATCTCTCAGCATCTTCTGATTTCTACAAGCCCA 600
|||||
601 TTGCTCTCAGCCTCTTCCAGACATCAGAGGGCAAAATACCTGCTGTCTGATCTGGGCCCA 660
|||||
601 TTGCTCTCAGCCTCTTCCAGACATCAGAGGGCAAAATACCTGCTGTCTGATCTGGGCCCA 660
|||||
661 CAAGCGTATGCAATCTCTAAGTTTACGGGACTCTCAATGACCACTATCAGTCACTGG 720
|||||
661 CAAGCGTATGCAATCTCTAAGTTTACGGGACTCTCAATGACCACTATCAGTCACTGG 720
|||||
721 CTGGCAACCTCAAGTACAGCTTAGGAAAACGGGGGACAAAAATTTCTGAAAAACATG 780
|||||
721 CTGGCAACCTCAAGTACAGCTTAGGAAAACGGGGGACAAAAATTTCTGAAAAACATG 780
|||||
781 GACAAAGGCCACCCCATCTTTTATTGAGTGAATGTGCTCTCCAGTTTCAGAACCCCTTCT 840
|||||
781 GACAAAGGCCACCCCATCTTTTATTGAGTGAATGTGCTCTCCAGTTTCAGAACCCCTTCT 840
|||||
841 ACCTACATCAGTCACTTAGATCTCAGCTGGGTTTCCAAATGAGAGCATGACCCGCTTG 900
|||||
841 ACCTACATCAGTCACTTAGATCTCAGCTGGGTTTCCAAATGAGAGCATGACCCGCTTG 900
|||||
901 TCAGTGGACCAAGCAAGGTGGAGCAAGAGATCTCCCGGTATCTCGGCTCAGAGG 960
|||||
901 TCAGTGGACCAAGCAAGGTGGAGCAAGAGATCTCCCGGTATCTCGGCTCAGAGG 960
|||||
961 TCTCCAGAAACAATAGTGTCCGGAAGGACACAGACTCTAAATTCAGTGTAAAGTTGTGC 1020
|||||
961 TCTCCAGAAACAATAGTGTCCGGAAGGACACAGACTCTAAATTCAGTGTAAAGTTGTGC 1020
|||||
1021 TGTCCGACATTTGTGACCAACATGCGGTAAACTCCACCTAAGCAAAAGCCAGCAAG 1080
|||||
1021 TGTCCGACATTTGTGACCAACATGCGGTAAACTCCACCTAAGCAAAAGCCAGCAAG 1080
|||||
1081 TCACCCGAACACCATTTACAGTTTGTAAACAGAGTGTGATGAAGAAATAGCTCTCAGAGCG 1140
|||||
1081 TCACCCGAACACCATTTACAGTTTGTAAACAGAGTGTGATGAAGAAATAGCTCTCAGAGCG 1140
|||||
1141 AATGCTCTAGTTTCCACTTTCCAGCCTGGATCCCTCACACTGAACCCCTTCTTCGTGCA 1200
|||||
1141 AATGCTCTAGTTTCCACTTTCCAGCCTGGATCCCTCACACTGAACCCCTTCTTCGTGCA 1200
|||||
1201 CCATCTGCTCTTGACATTTGAACTCAATGAACTCTCTGAGACCTGGCTCTGAGAAGA 1260
|||||
1201 CCATCTGCTCTTGACATTTGAACTCAATGAACTCTCTGAGACCTGGCTCTGAGAAGA 1260
|||||
1261 CTCGCAAAAAAATTC 1288
|||||
1261 CTCGCAAAAAAATTC 1288

BD085729 1288 bp DNA linear PAT 27-AUG-2002
LOCUS Genes from the 20q13 amplicon and their uses.
DEFINITION BD085729
ACCESSION BD085729
VERSION BD085729.1 GI:22631339
KEYWORDS JP 2001524802-A/5.
SOURCE synthetic construct
ORGANISM synthetic construct
REFERENCE 1 (bases 1 to 1288)
AUTHORS Gray,J.W., Collins,C.C., Hwang,S.I., Godfrey,T., Kowbel,D. and Rommens,J.
TITLE Genes from the 20q13 amplicon and their uses
JOURNAL Patent: JP 2001524802-A 5 04-DEC-2001;
COMMENT THE REGENTS OF THE UNIVERSITY OF CALIFORNIA
OS Artificial Sequence
PN JP 2001524802-A/5
PD 04-DEC-2001
PF 15-JUL-1997 JP 1998506264
PR 15-JUL-1996 US 08/680395,16-OCT-1996 US 08/731499 PR
PI 17-JAN-1997 US 08/785532
PI JOE W GRAY, COLIN CONRAD COLLINS, SOO IN HWANG, TONY GODFREY, PI DAVID KOWBEL,
PI JOHANNA ROMMENS
PC C12N15/11, C12Q1/68, A61K48/00
CC Description of Artificial Sequence:41.1
FH Key Location/Qualifiers
FT source 1..1288
FT /organism='Artificial Sequence'.
FEATURES
source
1..1288
/organism='synthetic construct'
/mol_type='genomic DNA'
/db_xref='taxon:32630'
ORIGIN
Query Match 100.0%; Score 1288; DB 6; Length 1288;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1288; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GAGGGCAGCGAAGAGAGAAACCCAGAGCCCTTGGAGCCCAATCTGCTCTGAGCAATGGG 60
DB 1 GAGGGCAGCGAAGAGAGAAACCCAGAGCCCTTGGAGCCCAATCTGCTCTGAGCAATGGG 60
QY 61 TCGGCTCTCGCAACCAAGCCCGCTGCGCATGATCAACCCACTCAGCGCCCTGCAAG 120
DB 61 TCGGCTCTCGCAACCAAGCCCGCTGCGCATGATCAACCCACTCAGCGCCCTGCAAG 120
QY 121 TCGGCTCTCGCAACCAATCACTTTGGGCAAGCCAGGAGCCCTTGGCTTCACTTCTGCTCC 180
DB 121 TCGGCTCTCGCAACCAATCACTTTGGGCAAGCCAGGAGCCCTTGGCTTCACTTCTGCTCC 180
QY 181 AGCCCAAGTTCAAGCACAATTTCCATGTTTCCACAAAGTCAATCTCAATGTCATGCAAG 240
DB 181 AGCCCAAGTTCAAGCACAATTTCCATGTTTCCACAAAGTCAATCTCAATGTCATGCAAG 240
QY 241 CCGGCTCTGAGTCTCGCTCCCAAGGTCAGCCAGCGGTCTCCAGGCGCTACCTGTTTGG 300
DB 241 CCGGCTCTGAGTCTCGCTCCCAAGGTCAGCCAGCGGTCTCCAGGCGCTACCTGTTTGG 300
QY 301 AACAGCGATCAGCCCATTTGACCTGACCAAGTCCAAAGCAAGAAAGCCGAGTCTTCGCAA 360
DB 301 AACAGCGATCAGCCCATTTGACCTGACCAAGTCCAAAGCAAGAAAGCCGAGTCTTCGCAA 360
QY 361 GCACAAATCTTGATGTCTCCACCTCAGAAAGCAGCTCTGTCTGACATCGCGACATGGTC 420
DB 361 GCACAAATCTTGATGTCTCCACCTCAGAAAGCAGCTCTGTCTGACATCGCGACATGGTC 420
QY 421 AAAGTCTCTCCAAAGCCACACCCCAAGCCAGCTCTCTCTCCAGGTCCTCCCTCCATG 480
DB 421 AAAGTCTCTCCAAAGCCACACCCCAAGCCAGCTCTCTCTCCAGGTCCTCCCTCCATG 480
QY 481 AAAGTGGAAATGATGTCTCAGGCGCTTTGAGGATGTCTCCAGTGAAGTCTCAACTTTGCAT 540


```
Db 481 AAGCTGGAATGGATGTCAGGCGCTTTGAGGATGCTCCAGTGAAGTCTCAACTTTGCAAT 540
Qy 541 AAAAGAAAAGCGCGCAGTCAACTCGGAATCTCAGCATCTTCTGATTTCTCAAGCCCGAG 600
Db 541 AAAAGAAAAGCGCGCAGTCCAACTCGGAATCTCAGCATCTTCTGATTTCTCAAGCCCGAG 600
Qy 601 TTTGCTCGAGCTTTCCAGACATCAGAGGCAAAATACCTGCTGCTGATCTGAGTCTGGGCCA 660
Db 601 TTTGCTCGAGCTTTCCAGACATCAGAGGCAAAATACCTGCTGCTGATCTGAGTCTGGGCCA 660
Qy 661 CAAGAGCGTATGCAATCTCTAAGTTTACGGACTCTCAATGACCACTATCAGTCACTGG 720
Db 661 CAAGAGCGTATGCAATCTCTAAGTTTACGGACTCTCAATGACCACTATCAGTCACTGG 720
Qy 721 CTGGCCAACTCACTAGTACCTAGTACCTGAGTACCTGAGTACCTGAGTACCTGAGTACCT 780
Db 721 CTGGCCAACTCACTAGTACCTAGTACCTGAGTACCTGAGTACCTGAGTACCTGAGTACCT 780
Qy 781 GACAAAGGCCACCCCATCTTTTATTCAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 840
Db 781 GACAAAGGCCACCCCATCTTTTATTCAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 840
Qy 841 ACCTACATCAGTCACTAGTACCTGAGTACCTGAGTACCTGAGTACCTGAGTACCTGAGT 900
Db 841 ACCTACATCAGTCACTAGTACCTGAGTACCTGAGTACCTGAGTACCTGAGTACCTGAGT 900
Qy 901 TCAGTGGACCAAGCAAGTGTGAGCAAGATCTCCCGGATCTCGTGGCTCAGAGG 960
Db 901 TCAGTGGACCAAGCAAGTGTGAGCAAGATCTCCCGGATCTCGTGGCTCAGAGG 960
Qy 961 TCTCCAGAAACAATAGTGTCCGAGAGGACAGACTCTAAATTCAGTGTGAGTGTGTC 1020
Db 961 TCTCCAGAAACAATAGTGTCCGAGAGGACAGACTCTAAATTCAGTGTGAGTGTGTC 1020
Qy 1021 TGTCCGACATTTGTGAGCAACATCGGTAAATCTCCACCTTAAGCAAAACGACAGCAAG 1080
Db 1021 TGTCCGACATTTGTGAGCAACATCGGTAAATCTCCACCTTAAGCAAAACGACAGCAAG 1080
Qy 1081 TCACCCGAAACCAATTCAGTGTGTAAACAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 1140
Db 1081 TCACCCGAAACCAATTCAGTGTGTAAACAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 1140
Qy 1141 AATGCTTAGTGTTCACATTTCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1200
Db 1141 AATGCTTAGTGTTCACATTTCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1200
Qy 1201 CCAATCTGCTTCTGACATTTGAACTCAATTTGAACTCAATTTGAACTCAATTTGAACTCA 1260
Db 1201 CCAATCTGCTTCTGACATTTGAACTCAATTTGAACTCAATTTGAACTCAATTTGAACTCA 1260
Qy 1261 CTGCCAAAAAATTTTCAATTTTCAATTTTCAATTTTCAATTTTCAATTTTCAATTTTCA 1288
Db 1261 CTGCCAAAAAATTTTCAATTTTCAATTTTCAATTTTCAATTTTCAATTTTCAATTTTCA 1288

RESULT 3
HSM808295 1830 bp mRNA linear PRI 30-AUG-2003
LOCUS Homo sapiens mRNA; cDNA DKFZp686K2480 (from clone DKFZp686K2480).
DEFINITION BX648148
ACCESSION BX648148.1 GI:34367307
VERSION 1
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1830)
AUTHORS Ansong, W., Krieger, S., Regier, T., Rittmuller, C., Schwager, B.,
Newes, H.W., Weill, B., Amd, C., Osanger, A., Fobo, G., Han, M. and
Wiemann, S.
CONSRMT The German Human cDNA Consortium
TITLE Direct Submission
```

```
JOURNAL Submitted (27-AUG-2003) MIPS, Ingolstaedter Landstr.1, D-85764
COMMENT Neuherberg, Germany
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
Heidelberg/Germany) within the cDNA sequencing consortium of the
German Genome Project.
This clone (DKFZp686K2480) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzd.de Further
information about the clone and the sequencing project is available
at http://mips.gsf.de/proj/cDNA/.
FEATURES
Location/Qualifiers
1..1830
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="DKFZp686K2480"
/tissue_type="human endometrium carcinoma cell line"
/clone_lib="686 (synonym: hlcc3). Vector pSport1_Sfi; host
DH10B; sites SfiIA + SfiIB"
/dev_stage="adult"
1819
polyA_site
ORIGIN
Query Match 99.6%; Score 1283.4; DB 9; Length 1830;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1284; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 GAGGGCAGGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 60
Db 300 GAGGGCAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 359
Qy 61 TGCGCCCTCGCCAAACCAACGCGCCCTCGCATGATCAACCCCACTCAGCGCCCTCGAG 120
Db 360 TGCGCCCTCGCCAAACCAACGCGCCCTCGCATGATCAACCCCACTCAGCGCCCTCGAG 419
Qy 121 TCGTCTGTAACATCACTTGGGCAAGCAACGAGCCCTTGGCGTCACTTCTCTGCTCC 180
Db 420 TCGTCTGTAACATCACTTGGGCAAGCAACGAGCCCTTGGCGTCACTTCTCTGCTCC 479
Qy 181 AGCCCAAGTTCAAGCACAATTTCCATGTTCCACAGTGAATCTCAATGTCAATGACCAAG 240
Db 480 AGCCCAAGTTCAAGCACAATTTCCATGTTCCACAGTGAATCTCAATGTCAATGACCAAG 539
Qy 241 CGGTCTTGTAGTCTGCTCCACAGAGTCAAGCGGTGTCAGCGGTGTCAGCGGTGTCAG 300
Db 540 CGGTCTTGTAGTCTGCTCCACAGAGTCAAGCGGTGTCAGCGGTGTCAGCGGTGTCAG 599
Qy 301 AACAGCGATCAGCCCATTTGACCTGACCAAGTCCAAAGCAAGAAAGCGAGTCTCGCA 360
Db 600 AACAGCGATCAGCCCATTTGACCTGACCAAGTCCAAAGCAAGAAAGCGAGTCTCGCA 659
Qy 361 GCACATCTTGTATGTCCTCCACCTCAGAGCAAGCTGCTGTCGATCGCGACATGGTC 420
Db 660 GCACATCTTGTATGTCCTCCACCTCAGAGCAAGCTGCTGTCGATCGCGACATGGTC 719
Qy 421 AAAGTCTCTCCCAAGCAACCAACCCCAAGCAGCTCTCTCTCCAGGGTCTCCCGCCCATG 480
Db 720 AAAGTCTCTCCCAAGCAACCAACCCCAAGCAGCTCTCTCTCCAGGGTCTCCCGCCCATG 779
Qy 481 AAGCTGGAATTTGAGTGTTCAGGCGCTTTGAGGATGCTCCAGTGAAGTCTCAACTTTGCA 540
Db 780 AAGCTGGAATTTGAGTGTTCAGGCGCTTTGAGGATGCTCCAGTGAAGTCTCAACTTTGCA 839
Qy 541 AAAAGAAAAGCGCGCAGTCCAACTCGGAATCTCAGCATCTTCTGATTTCTCAAGCCCGAG 600
Db 840 AAAAGAAAAGCGCGCAGTCCAACTCGGAATCTCAGCATCTTCTGATTTCTCAAGCCCGAG 899
Qy 601 TTTGCTCGAGCTTTCCAGACATCAGAGGCAAAATACCTGCTGCTGATCTGAGTCTGGGCCA 660
Db 900 TTTGCTCGAGCTTTCCAGACATCAGAGGCAAAATACCTGCTGCTGATCTGAGTCTGGGCCA 959
```

```
QY 661 CAAGAGCGTATGCAAAATCTCTAAGTTTACGGGACTCTCAATGACCACTATCAGTCACTGG 720
Db 960 CAAGAGCGTATGCAAAATCTCTAAGTTTACGGGACTCTCAATGACCACTATCAGTCACTGG 1019
QY 721 CTGGCCAAACGTCAAGTACACAGCTTAGGAAACCGGGGGGACAAAATTTCTGAAAAACATG 780
Db 1020 CTGGCCAAACGTCAAGTACACAGCTTAGGAAACCGGGGGGACAAAATTTCTGAAAAACATG 1079
QY 781 GACAAAGGCCACCCCACTTTTATTCAGTGAAGTGTGCGCTCCAGTTTCAGAACCCCTTCT 840
Db 1080 GACAAAGGCCACCCCACTTTTATTCAGTGAAGTGTGCGCTCCAGTTTCAGAACCCCTTCT 1139
QY 841 ACCTACATCAGTCACCTAGAAATCTACCTGGGTTTCCAAATGAAGGACATGACCGCTTG 900
Db 1140 ACCTACATCAGTCACCTAGAAATCTACCTGGGTTTCCAAATGAAGGACATGACCGCTTG 1199
QY 901 TCAGTGGACCAAGCAAGTGGAGCAAGAGATCTCCGGGTATCGTCGGCTCAGAGG 960
Db 1200 TCAGTGGACCAAGCAAGTGGAGCAAGAGATCTCCGGGTATCGTCGGCTCAGAGG 1259
QY 961 TCTCCAGAAACAATAGTCGCGAAGAGGACACAGACTCTAAATTCAGGTGAAGTTGTGC 1020
Db 1260 TCTCCAGAAACAATAGTCGCGAAGAGGACACAGACTCTAAATTCAGGTGAAGTTGTGC 1319
QY 1021 TGTCCGACATTTGTGAGCAAAATGCGGTAAAACTCCACCTAAGCAAAAACGCACAGCAAG 1080
Db 1320 TGTCCGACATTTGTGAGCAAAATGCGGTAAAACTCCACCTAAGCAAAAACGCACAGCAAG 1379
QY 1081 TCACCCGAACACCATTCACAGTTTGTAAACAGAGCTGGATGAAGATAGCTCTGCGAGGAG 1140
Db 1380 TCACCCGAACACCATTCACAGTTTGTAAACAGAGCTGGATGAAGATAGCTCTGCGAGGAG 1439
QY 1141 AATGCGCTTAGTTTCCAGCTTTCAGAGCTGGATCCCTCAGACTGAAACCCCTTCTCGTTGCA 1200
Db 1440 AATGCGCTTAGTTTCCAGCTTTCAGAGCTGGATCCCTCAGACTGAAACCCCTTCTCGTTGCA 1499
QY 1201 CCATCTGCTTTCAGACTGAACTCAATGAACTCCCTCAGACCCCTGCTGAGAAGA 1260
Db 1500 CCATCTGCTTTCAGACTGAACTCAATGAACTCCCTCAGACCCCTGCTGAGAAGA 1559
QY 1261 CTGCCAAAAAATAAAAAAAAAAAAA 1285
Db 1560 CTGCCAAAAAATAAAAAAAAAAAAA 1584

RESULT 4
AX746848
LOCUS AX746848
DEFINITION Sequence 373 from Patent EP1308459.
ACCESSION AX746848
VERSION AX746848.1 GI:32131236
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1
AUTHORS Isogai,T., Sugiyama,T., Otsuki,T., Wakamatsu,A., Sato,H., Ishii,S.,
Yamamoto,J.I., Isono,Y., Hio,Y., Otsuka,K., Nagai,K., Irie,R.,
Tamechika,I., Seki,N., Yoshikawa,T., Otsuka,M., Nagahari,K. and
Masuho,Y.
TITLE Full-length cDNA sequences
JOURNAL Patent: EP 1308459-A 373 07-MAY-2003;
Biotechnology (JP) ; Research Association for
Helix Research Institute (JP)
FEATURES
source
Location/Qualifiers
1..3322
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
ORIGIN
Query Match 98.2%; Score 1265; DB 6; Length 3322;
```

```
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1265; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GAGGGCAGCGAAGAGAGAAACCCAGCCCCCTGGAGGCCACATCTGCTCTGAGCAATGGG 60
Db 2058 GAGGGCAGCGAAGAGAGAAACCCAGCCCCCTGGAGGCCACATCTGCTCTGAGCAATGGG 2117
QY 61 TCGCGCCTCGCCAAACCCAGCCCCCGGCTGCATGATCAACCCACTCAGCGCCCTGCAG 120
Db 2118 TCGCGCCTCGCCAAACCCAGCCCCCGGCTGCATGATCAACCCACTCAGCGCCCTGCAG 2177
QY 121 TCCGTCCTGAAACAATCACTTGGGCAAGCCAGGAGCCCTTGGCTCACCCTTCTGCTCC 180
Db 2178 TCCGTCCTGAAACAATCACTTGGGCAAGCCAGGAGCCCTTGGCTCACCCTTCTGCTCC 2237
QY 181 AGCCCAAGTTCAAGCACAAATTTCCATGTTCCACAAGTCAATCTCAATGTCATGGAAGA 240
Db 2238 AGCCCAAGTTCAAGCACAAATTTCCATGTTCCACAAGTCAATCTCAATGTCATGGAAGA 2297
QY 241 CCGGTCCTTCCAGTCCCTCCAAAGGTCAGCCAGCGTCTCCAGGCGCTACCTGTTTGAG 300
Db 2298 CCGGTCCTTCCAGTCCCTCCAAAGGTCAGCCAGCGTCTCCAGGCGCTACCTGTTTGAG 2357
QY 301 AACAGCGATCAGCCCAATTCACCTGACCAAGTCCAAAAGCAAGAAAGCCGAGTCTCTCGCAA 360
Db 2358 AACAGCGATCAGCCCAATTCACCTGACCAAGTCCAAAAGCAAGAAAGCCGAGTCTCTCGCAA 2417
QY 361 GCACAATCTTGTATGTGCCCACTCAGAAAGCAGCTGTCTGACATCGCCGACATGGTC 420
Db 2418 GCACAATCTTGTATGTGCCCACTCAGAAAGCAGCTGTCTGACATCGCCGACATGGTC 2477
QY 421 AAAGTCCTCCCAAGAGCCACCAAGCCAGCTCTCTCCAGGCGTCCCCCCCCATG 480
Db 2478 AAAGTCCTCCCAAGAGCCACCAAGCCAGCTCTCTCCAGGCGTCCCCCCCCATG 2537
QY 481 AAGCTGGAAATGGATGTTCAGGCGCTTTGAGGATGTCTCCAGTCAAGTCTCAACTTTGCAT 540
Db 2538 AAGCTGGAAATGGATGTTCAGGCGCTTTGAGGATGTCTCCAGTCAAGTCTCAACTTTGCAT 2597
QY 541 AAAAGAAAAGGCGCGGAGTCCAACTGGAACTCTCAGCATCTTCTGATTTACAGGCCAG 600
Db 2598 AAAAGAAAAGGCGCGGAGTCCAACTGGAACTCTCAGCATCTTCTGATTTACAGGCCAG 2657
QY 601 TTTTGCTCGAGCTCTTCCAGACATCAGAGGGCAAAATACCTGCTGCTCATCTGGGCCCA 660
Db 2658 TTTTGCTCGAGCTCTTCCAGACATCAGAGGGCAAAATACCTGCTGCTCATCTGGGCCCA 2717
QY 661 CAAGAGCGTATGCAAAATCTCTAAGTTTACGGGACTCTCAATGACCACTATCAGTCACTGG 720
Db 2718 CAAGAGCGTATGCAAAATCTCTAAGTTTACGGGACTCTCAATGACCACTATCAGTCACTGG 2777
QY 721 CTGGCCAAACGTCAAGTACCAGCTTAGGAAAACGGGGGGGACAAAATTTCTGAAAAACATG 780
Db 2778 CTGGCCAAACGTCAAGTACCAGCTTAGGAAAACGGGGGGGACAAAATTTCTGAAAAACATG 2837
QY 781 GACAAAGGCCACCCCACTTTTATTCAGTGAAGTGTGCGCTCCAGTTTCAGAACCCCTTCT 840
Db 2838 GACAAAGGCCACCCCACTTTTATTCAGTGAAGTGTGCGCTCCAGTTTCAGAACCCCTTCT 2897
QY 841 ACCTACATCAGTCACCTAGAAATCTCACCCTGGGTTTCCAAATGAAGGACATGACCGCTTG 900
Db 2898 ACCTACATCAGTCACCTAGAAATCTCACCCTGGGTTTCCAAATGAAGGACATGACCGCTTG 2957
QY 901 TCAGTGGACCAAGCAAGTGGAGCAAGAGATCTCCGGGTATCGTCGGCTCAGAGG 960
Db 2958 TCAGTGGACCAAGCAAGTGGAGCAAGAGATCTCCGGGTATCGTCGGCTCAGAGG 3017
QY 961 TCTCCAGAAACAATAGCTGCCGAAGAGGACACAGACTCTAAATTCAGGTGAAGTTGTGC 1020
Db 3018 TCTCCAGAAACAATAGCTGCCGAAGAGGACACAGACTCTAAATTCAGGTGAAGTTGTGC 3077
QY 1021 TCTCCGACATTTGTGAGCAAAACATCGGTTAAACTCCACTAAGCAAAAACGCACAGCAAG 1080
Db 1080 TCTCCGACATTTGTGAGCAAAACATCGGTTAAACTCCACTAAGCAAAAACGCACAGCAAG 1080
```

Db	3078	TGTCGGACATTGTGTGAGCAACATCGCGGTAAACCTCCACCTAAGCAAAACGCACAGCAAG	3137
Qy	1081	TCACCCGAACACCATTCACAGTTTCTAACAGACGTGGATGAAGATAGCTCTGCGAGGACG	1140
Db	3138	TCACCCGAACACCATTCACAGTTTCTAACAGACGTGGATGAAGATAGCTCTGCGAGGACG	3197
Qy	1141	AATGCTTAGTTTTCACATTTCCAGCTCGATCCCTACACCTGACCTGACCTCTTCTGTTGCA	1200
Db	3198	AATGCTTAGTTTTCACATTTCCAGCTCGATCCCTACACCTGACCTGACCTCTTCTGTTGCA	3257
Qy	1201	CCATCTCTGCTCTGACATTTGAACCTCATTGAACCTCTCTGACACCTCGCTCTGAGAAGA	1260
Db	3258	CCATCTCTGCTCTGACATTTGAACCTCATTGAACCTCTCTGACACCTCGCTCTGAGAAGA	3317
Qy	1261	CTGCC 1265	
Db	3318	CTGCC 3322	
RESULT 5			
AK091206			
LOCUS			
DEFINITION	AK091206	3322 bp mRNA linear	PRI 30-JAN-2004
ACCESSION	AK091206	hom sapiens cdna FLJ33887 fis, clone CRONG2007400, highly similar	
VERSION	AK091206.1	GI:21749522	
KEYWORDS		oligo capping; fis (full insert sequence).	
SOURCE		Hom sapiens (human)	
ORGANISM		Hom sapiens	
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
AUTHORS		Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
		Ota, T., Suzuki, Y., Nishikawa, T., Otsuki, T., Sugiyama, T., Irie, R.,	
		Wakamatsu, A., Hayashi, K., Sato, H., Nagai, K., Kimura, K., Makita, H.,	
		Sekine, M., Obayashi, M., Nishi, T., Shibahara, T., Tanaka, T.,	
		Ishii, S., Yamamoto, J., Saito, K., Kawai, Y., Isono, Y., Nakamura, Y.,	
		Nagahari, K., Murakami, K., Yasuda, T., Iwayanagi, T., Wagatsuma, M.,	
		Shiratori, A., Sudo, H., Hosoiri, T., Kaku, Y., Kodaira, H., Kondo, H.,	
		Sugawara, M., Takahashi, M., Kanda, K., Yokoi, T., Furuya, T.,	
		Kikkawa, E., Omura, Y., Abe, K., Kamihara, K., Katsuta, N., Sato, K.,	
		Tanikawa, M., Yamazaki, M., Ninomiya, K., Ishibashi, T., Yamashita, H.,	
		Murakawa, K., Fujimori, K., Tanai, H., Kimata, M., Watanabe, M.,	
		Hiraoka, S., Chiba, Y., Ishida, S., Ono, Y., Takiguchi, S., Watanabe, S.,	
		Yosida, M., Hotuta, T., Kusano, J., Kanehori, K., Takahashi, Fujii, A.,	
		Hara, H., Tanase, T., Nomura, Y., Togiya, S., Komai, F., Hara, R.,	
		Takeuchi, K., Arita, M., Imose, N., Musashino, K., Yuuki, H., Oshima, A.,	
		Sasaki, N., Aotsuka, S., Yoshikawa, Y., Matsunawa, H., Ichihara, T.,	
		Shichata, N., Sano, S., Moriya, S., Momiyama, H., Satoh, N., Takami, S.,	
		Terashima, Y., Suzuki, O., Nakagawa, S., Senoh, A., Mizoguchi, H.,	
		Goto, Y., Shimizu, F., Wakebe, H., Hishigaki, H., Watanabe, T.,	
		Sugiyama, A., Takemoto, M., Kawakami, B., Yamazaki, M., Watanabe, K.,	
		Kumagai, A., Itakura, S., Fukuzumi, Y., Fujimori, Y., Komiyama, M.,	
		Takahiro, H., Tanigami, A., Fujiwara, T., Ono, T., Yamada, K., Fujii, Y.,	
		Ozaki, K., Hirao, M., Ohnori, Y., Kawabata, A., Hikiji, T., Kobatake, N.,	
		Inguchi, H., Ikema, Y., Okamoto, S., Okitani, R., Kawakami, T.,	
		Noguchi, S., Itoh, T., Shigetani, K., Senba, T., Matsumura, K.,	
		Nakajima, Y., Mizuno, T., Morinaga, M., Sasaki, M., Togaishi, T.,	
		Oyama, M., Hata, H., Watanabe, M., Komatsu, T., Mizushima-Sugano, J.,	
		Satoh, T., Shirai, Y., Takahashi, Y., Nakagawa, K., Okumura, K.,	
		Nagase, T., Nomura, N., Kikuchi, H., Masuho, Y., Yamashita, R.,	
		Nakai, K., Yada, T., Nakamura, Y., Ohara, O., Isogai, T. and Sugano, S.	
		Complete sequencing and characterization of 21,243 full-length	
		human cDNAs	
		Nat. Genet. 36 (1), 40-45 (2004)	
		14702039	
		2	
REFERENCE			
AUTHORS			
		Nishi, T., Ota, T., Nakagawa, S., Senoh, A., Mizuguchi, H., Inagaki, H.,	
		Sugiyama, T., Irie, R., Otsuki, T., Sato, H., Wakamatsu, A., Ishii, S.,	
		Yamamoto, J., Isono, Y., Kawai-Hio, Y., Saito, K., Nishikawa, T.,	
		Kimura, K., Yamashita, H., Matsuo, K., Nakamura, Y., Sekine, M.,	
		Kikuchi, H., Kanda, K., Wagatsuma, M., Murakawa, K., Kanehori, K.,	
		Takahashi-Fujii, A., Oshima, A., Sugiyama, A., Kawakami, B., Suzuki, Y.,	
		Sugano, S., Nagahari, K., Masuho, Y., Nagai, K. and Isogai, T.	

TITLE	NEDO human cDNA sequencing project									
JOURNAL	Unpublished									
REFERENCE	3 (bases 1 to 3322)									
AUTHORS	Isogai, T. and Yamamoto, J.									
TITLE	Direct Submission									
JOURNAL	Submitted (04-JUL-2002) Takao Isogai, FLJ Project (HRI Team); 2-6-7									
COMMENT	Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan (E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986) NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5' - & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: HRI and RAB; annotation: HRI and RAB.									
FEATURES	Location/Qualifiers									
source	1..3322 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="CTONG2007400" /tissue_type="tongue, tumor tissue" /clone_lib="CTONG2" /note="cloning vector: pME18SFL3" 81..3318 /note="unnamed protein product" /codon_start=1 /protein_id="BAC03610.1" /db_xref="GI:21749523" /translation="MPRRKQAPRAAGYAQEBQLKEEELKEEEDSDSSVAQLQGD NDTGTDELETGEQCFYSQNSGSHLSNQDAENESLLSDSDQVSDIKSVCGRD ASDKKATHVRLPNEAHNCMDKMTAVANTILSDSYWSGLGFLKLSNERNCDTRNG SNKDFDHWDLALSKQNLPSRSVSKPSLFSVQVLYRQSSKCGTFTGASRPRCR QCSAAVDTLVELTVHNETGHYQDDNKKDKLRPTYSKPKRAFDQMDKEDAKVLR CMFGDSDLSQDLSVMIKTKYQVPLKEPVTISKMYTPAKKRVDFVDRPCSPD STTSGPADSPSSQKNAHLQSSNNRYQNGASTWQFEACKSQILKMEGSSHDTL QDLTHMMVTGHFLUKTSASAKKQQLVDPLAVKQSLSEANSDSLAPKPSNSA SDCTASTTELKESKEPPEETSKDEKVSIEDYEDLPQKPLDTIKYLYREDEEDL GSKGGDILKSLNTVTTAINAGNAPSPWSAYPSIAHAYQLSGEKPEPLPMGRVLO IRPNLTNKLRIAPKWPVLVSMPTLAPYTVQVKESDEKDEAKVEKCEGSPHEAS SPFSGSDSPKSTPEAKTELGLKEEKLMEGSEKPEKPPLEPSTALSNGCAL ANHPALPCLNPLSALQSLNHLGKATEPLRSPSCSPSSSTISMFHKNLNVMDKP VLSPATRSASVRRYLFENSQDPIDLTKSKKAESSQAQSCWSPKPHALSDIADM VKVLPAATTPKPASSRVPKMLKEDVRPDEVSEVSTLHKRGRQSNWNPQHLIL QAQPASSLFTSEGKYLISDLGPQRMQISKFTGLSMTTISHLANVYQLRKTGTTK FLKNMDKGRHPIFYCSDCASQFPTPTVISHLESHLGFQMDKTRLSDVDQSKVEQEIL RVSSAQRSPETIAEEDTDSKFKCLCRFTVSKHAVKHLHLSKTHSKSPHEHSQFVTD VDEE"									
CDS										
ORIGIN										
Query Match	98.2%; Score 1265; DB 9; Length 3322;									
Best Local Similarity	100.0%; Pred. No. 0;									
Matches 1265; Conservative	0; Mismatches 0; Indels 0; Gaps 0;									
Qy	1	GAGGGCAGCAGAGAGAGAAACCCAGCCCTCGAGGCCACATCTGCTCTGAGCAATGGG	60							
Db	2058	GAGGGCAGCAGAGAGAGAAACCCAGCCCTCGAGGCCACATCTGCTCTGAGCAATGGG	2117							
Qy	61	TGCGCCCTCGCACAACCGCCCGCCCTGCGCATGCATCAACCCACTCAGCGCCCTGCAG	120							
Db	2118	TGCGCCCTCGCACAACCGCCCGCCCTGCGCATGCATCAACCCACTCAGCGCCCTGCAG	2177							
Qy	121	TCCGCTCTGAAACAATCACTTGGGCAAGCAGGAGCCCTTGGCTTCACTTCTCTCTCC	180							
Db	2178	TCCGCTCTGAAACAATCACTTGGGCAAGCAGGAGCCCTTGGCTTCACTTCTCTCTCC	2237							
Qy	181	AGCCCAAGTTCAAGCACAAATTTCCATGTTCCCAAGTCGAATCTCAATGTCTATGACAAAG	240							
Db	2238	AGCCCAAGTTCAAGCACAAATTTCCATGTTCCCAAGTCGAATCTCAATGTCTATGACAAAG	2297							
Qy	241	CCGGTCTTGAGTCTGCTTCCCAAGGTCACGCGGTGTCCAGGCGCTACCTGTTTGAG	300							

2298	DB	CCGGTCTTGAGTCTCTGCCTCCACAAGGTCTAGCCAGCGGTGTCTCAGGGCGCTACCTGTTTGAG	2355
301	QY	AAACAGCGATCAGCCCATTTGACTCTGACCAAGTCCAAAAGCAAGAAAGCCGAGTCTCTCGCAA	360
2358	DB	AAACAGCGATCAGCCCATTTGACTCTGACCAAGTCCAAAAGCAAGAAAGCCGAGTCTCTCGCAA	2417
361	QY	GCACAAATCTTTGTATGTCTCCCACTCTCAGAAGGACGCTCTGTCTGACATCGCGCGACATGGTC	420
2418	DB	GCACAAATCTTTGTATGTCTCCCACTCTCAGAAGGACGCTCTGTCTGACATCGCGCGACATGGTC	2477
421	QY	AAAGTCTCTCCCAAGGACACACCCAAAAGCGACGCTCTCTTCAGGGTCCCGCCCATG	480
2478	DB	AAAGTCTCTCCCAAGGACACACCCAAAAGCGACGCTCTCTTCAGGGTCCCGCCCATG	2537
481	QY	AAAGTGGAAATGGATGTCTCAGCGCTTTGAGGATGTCTCCAGTGAAGTCTCAACTTTGTCAT	540
2538	DB	AAAGTGGAAATGGATGTCTCAGCGCTTTGAGGATGTCTCCAGTGAAGTCTCAACTTTGTCAT	2597
541	QY	AAAAGAAAGGCGCGGAGTCCAACTGGAATCCTCAGCATCTCTGATTTCTACAAAGCCCGAG	600
2598	DB	AAAAGAAAGGCGCGGAGTCCAACTGGAATCCTCAGCATCTCTGATTTCTACAAAGCCCGAG	2657
601	QY	TTTGCCTCTGAGCCCTCTTCCAGACATCAGAGGGCAAAATACCTGCTGTCTGATCTGGGCCCA	660
2658	DB	TTTGCCTCTGAGCCCTCTTCCAGACATCAGAGGGCAAAATACCTGCTGTCTGATCTGGGCCCA	2717
661	QY	CAAGAGCGTATGAAATCTCTAAGTTTACGGGACTCTCAATAGACCACTATCAGTCACTGG	720
2718	DB	CAAGAGCGTATGAAATCTCTAAGTTTACGGGACTCTCAATAGACCACTATCAGTCACTGG	2777
721	QY	CTGSCCAACGTCGAGTACCACTTTAGGAAACGGGCGGGAACAAATTTCTGAAAAACATG	780
2778	DB	CTGSCCAACGTCGAGTACCACTTTAGGAAACGGGCGGGAACAAATTTCTGAAAAACATG	2837
781	QY	GACAAAGGCCACCCCATCTTTATTGTCAGTGACTGTGCCTCCAGTTTCAGAACCCCTTCT	840
2838	DB	GACAAAGGCCACCCCATCTTTATTGTCAGTGACTGTGCCTCCAGTTTCAGAACCCCTTCT	2897
841	QY	ACCTACATCAGTCACTTAGAATCTCAGCTGGGTTTCCAAATGAAGACATGACCCGCTTG	900
2898	DB	ACCTACATCAGTCACTTAGAATCTCAGCTGGGTTTCCAAATGAAGACATGACCCGCTTG	2957
901	QY	TCAGTGGACAGCAAAAGCGAGGTGGAGCAAGAGATCTCCGGGTATCGTCGGCTCAGAGG	960
2958	DB	TCAGTGGACAGCAAAAGCGAGGTGGAGCAAGAGATCTCCGGGTATCGTCGGCTCAGAGG	3017
961	QY	TCTCCAGAAACAATAGCTGCCGGAAGAGGACACAGACTCTAAATTCAGTGTAGTTGTGC	1020
3018	DB	TCTCCAGAAACAATAGCTGCCGGAAGAGGACACAGACTCTAAATTCAGTGTAGTTGTGC	3077
1021	QY	TGTCGGACATTTGTGAGCAAAACATGCGGTAAACTCCACCTTAAGCAAAAACGACAGCAAG	1080
3078	DB	TGTCGGACATTTGTGAGCAAAACATGCGGTAAACTCCACCTTAAGCAAAAACGACAGCAAG	3137
1081	QY	TCACCCGAAACCAATTCACAGTTGTTTAAACAGACGTGGATGAAGAAATAGCTCTGCAGGACG	1140
3138	DB	TCACCCGAAACCAATTCACAGTTGTTTAAACAGACGTGGATGAAGAAATAGCTCTGCAGGACG	3197
1141	QY	AATGCCCTTAGTTTCCACTTTCAGCGCTGGATCCCTCACACTGAACCCCTTCTGCTTGCA	1200
3198	DB	AATGCCCTTAGTTTCCACTTTCAGCGCTGGATCCCTCACACTGAACCCCTTCTGCTTGCA	3257
1201	QY	CCATCTCTGCTCTGACATTTGAATCTCATTTGAATCCTCTGACACCCCTGGCTCTGAGAAGA	1260
3258	DB	CCATCTCTGCTCTGACATTTGAATCTCATTTGAATCCTCTGACACCCCTGGCTCTGAGAAGA	3317
1261	QY	CTGCC	1265
3318	DB	CTGCC	3322

LOCUS	EX276189	54161 bp	DNA	linear	PRI 25-FEB-2003
DEFINITION	Human DNA sequence from clone RP4-678D15 on chromosome 20q13.13-13.2, complete sequence.				
ACCESSION	EX276189				
VERSION	EX276189.1	GI:28564238			
KEYWORDS	HTG.				
SOURCE	Homo sapiens	(human)			
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	1 (bases 1 to 54161)				
TITLE	Matthews, L.				
JOURNAL	Direct Submission				
	Submitted (25-FEB-2003) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk				
COMMENT	----- Genome Center Center: Wellcome Trust Sanger Institute Center code: SC Web site: http://www.sanger.ac.uk Contact: humquery@sanger.ac.uk				

OPTGTN

	Query Match	88.3%;	Score 1137;	DB 9;	Length 54161;
	Best Local Similarity	100.0%;	Pred. No. 8e-310;	Mismatches 0;	Gaps 0;
	Matches 1137;	Conservative	0;	Indels	0;
Qy	1	GAGGGCAGCGAAGGAGAAACCCAGGCCCTTGGAGCCCATCTGCCTCTCAGCAATGGG	60		
Dd	40512	GAGGGCAGCGAAGGAGAAACCCAGGCCCTTGGAGCCCATCTGCCTCTCAGCAATGGG	40571		
Qy	61	TGGGCCCTTCGCAAACACAGCCCGGCCCTTGCCATCGATCAACCCTCAGCGCCCTTGAG	120		
Dd	40572	TGGGCCCTTCGCAAACACAGCCCGGCCCTTGCCATCGATCAACCCTCAGCGCCCTTGAG	40631		
Qy	121	TCGCGTCTGAACAACACTTTGGGCGAAAGCCAGCGCCCTTTCGCTCACCTTCCTGTGCTCC	180		

```
Db 40632 TCCGTCCTGAAACAATCACTTGGGCAAGACCGAGGCCCTTGGCGTCACTTCTGCTGCC 40691
QY 181 AGCCCAAGTTCAAGCACAAATTTCCATGTTTCCACAAGTGAATCTCAATGTCTATGACACAAG 240
Db 40692 AGCCCAAGTTCAAGCACAAATTTCCATGTTTCCACAAGTGAATCTCAATGTCTATGACACAAG 40751
QY 241 CCGGTCCTTGAAGTCTGCTCCACAAAGGTCAGCAGCGGTGTCCAGCGGTCTACCTGTTTGG 300
Db 40752 CCGGTCCTTGAAGTCTGCTCCACAAAGGTCAGCAGCGGTGTCCAGCGGTCTACCTGTTTGG 40811
QY 301 AACAGCGATCAGCCCAATGACTGACCAAGTCCAAAGCAAGACGAGTCTCTCGCAA 360
Db 40812 AACAGCGATCAGCCCAATGACTGACCAAGTCCAAAGCAAGACGAGTCTCTCGCAA 40871
QY 361 GCACAATCTTGTATGTCCTCCACCTCAGAGCACGCTCTGTCTGTGACATCGCCGACATGTC 420
Db 40872 GCACAATCTTGTATGTCCTCCACCTCAGAGCACGCTCTGTCTGTGACATCGCCGACATGTC 40931
QY 421 AAAGTCTCTCCCAAGGACACACCCCAAGGAGCCTCTCTCTCCAGGGTCCGCCCATG 480
Db 40932 AAAGTCTCTCCCAAGGACACACCCCAAGGAGCCTCTCTCTCCAGGGTCCGCCCATG 40991
QY 481 AAGCTGGAATGGATGTCAGGCGCTTTCAGGATGTCCTCAGTGAAGTCTCAACTTTCAT 540
Db 40992 AAGCTGGAATGGATGTCAGGCGCTTTCAGGATGTCCTCAGTGAAGTCTCAACTTTCAT 41051
QY 541 AAAAGAAAAGGCGGACGTCACAACTGGAATCTCAGCATCTTCTGATTTCTCAAGACCCAG 600
Db 41052 AAAAGAAAAGGCGGACGTCACAACTGGAATCTCAGCATCTTCTGATTTCTCAAGACCCAG 41111
QY 601 TTTGCTCTGAGCCTCTCCAGACATCAGAGGCAAACTCTGCTGTCTGATCTCGGCCCA 660
Db 41112 TTTGCTCTGAGCCTCTCCAGACATCAGAGGCAAACTCTGCTGTCTGATCTCGGCCCA 41171
QY 661 CAAGAGCGTATGCAATCTTAAGTTTACGGACTCTCAATGACCACTATCAGTCACTGG 720
Db 41172 CAAGAGCGTATGCAATCTTAAGTTTACGGACTCTCAATGACCACTATCAGTCACTGG 41231
QY 721 CTGGCCAAAGTCAAGTACCAAGTACCAAGTACCAAGTACCAAGTACCAAGTACCAAGTAC 780
Db 41232 CTGGCCAAAGTCAAGTACCAAGTACCAAGTACCAAGTACCAAGTACCAAGTACCAAGTAC 41291
QY 781 GACAAAGGCCACCCCATCTTTTATGTCAGTGAAGTGTGCTCTCCAGTTCAGAACCCCTTCT 840
Db 41292 GACAAAGGCCACCCCATCTTTTATGTCAGTGAAGTGTGCTCTCCAGTTCAGAACCCCTTCT 41351
QY 841 ACCTACATCAGTCACTTAGAATCTCACTGGGTTTCCAAATGAGGACATGACCGCTTG 900
Db 41352 ACCTACATCAGTCACTTAGAATCTCACTGGGTTTCCAAATGAGGACATGACCGCTTG 41411
QY 901 TCAGTGGACCAAGCAAGTGGAGCAAGATCTCCCGGTTATCGTGGCTCAGAGG 960
Db 41412 TCAGTGGACCAAGCAAGTGGAGCAAGATCTCCCGGTTATCGTGGCTCAGAGG 41471
QY 961 TCTCCAGAAACAATAGCTGCGGAGAGGACAGACACTCTAAATCTCAAGTGAAGTGTGTC 1020
Db 41472 TCTCCAGAAACAATAGCTGCGGAGAGGACAGACACTCTAAATCTCAAGTGAAGTGTGTC 41531
QY 1021 TGTGGACATTTGTGAGCAAAATCGGTAAATCTCCACCTTAAGCAAAACGACAGCAAG 1080
Db 41532 TGTGGACATTTGTGAGCAAAATCGGTAAATCTCCACCTTAAGCAAAACGACAGCAAG 41591
QY 1081 TCACCCGAACACCATTTACAGTTTCTTAACAGACGTTGAGATGAGATAGCTCTGCGAGG 1137
Db 41592 TCACCCGAACACCATTTACAGTTTCTTAACAGACGTTGAGATGAGATAGCTCTGCGAGG 41648
```

RESULT 7

AC051637/c

LOCUS

DEFINITION

SEQUENCE, 27 unordered pieces.

172256 bp DNA linear HTG 24-AUG-2002
Homo sapiens chromosome 20 clone RP11-664G18 map 20, WORKING DRAFT
SEQUENCE, 27 unordered pieces.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS
TITLE
JOURNAL
AUTHORS

TITLE
JOURNAL
COMMENT

AC051637.2 GI:7770475
HTG; HTGS PHASE1; HTGS_DRAFT.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 172256)
Birren, B., Linton, L., Nusbaum, C. and Lander, E.
Homo sapiens chromosome 20, clone RP11-664G18
Unpublished

2 (bases 1 to 172256)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F.,
Boguslavskiy, L., Boukhgalter, B., Brown, A., Burkett, G.,
Campiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S.,
Collamore, A., Cooke, P., DeArelano, K., Dewar, K., Diaz, J.S.,
Dodge, S., Domino, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D.,
Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,
Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,
Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
Klein, J., LaRocque, K., Lamazares, R., Landers, T., Lehoczy, J.,
Levine, R., Lieu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N.,
McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R.,
Meldrim, J., Meneus, L., Mihova, T., Miranda, C., Mlenga, V., Morrow, J.,
Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
O'Neil, D., Oliver, T.M., Oliver, J., Peterson, K., Pierre, N.,
Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Teafaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J.,
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,
Young, G., Zainoun, J., Zimmer, A. and Zody, M.

Direct Submission
Submitted (15-APR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA

3 (bases 1 to 172256)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F.,
Boguslavskiy, L., Boukhgalter, B., Brown, A., Burkett, G.,
Campiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S.,
Collamore, A., Cooke, P., DeArelano, K., Dewar, K., Diaz, J.S., Gage, D.,
Dodge, S., Domino, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D.,
Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,
Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,
Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
Klein, J., LaRocque, K., Lamazares, R., Landers, T., Lehoczy, J.,
Levine, R., Lieu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N.,
McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R.,
Meldrim, J., Meneus, L., Mihova, T., Miranda, C., Mlenga, V., Morrow, J.,
Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
O'Neil, D., Oliver, T.M., Oliver, J., Peterson, K., Pierre, N.,
Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Teafaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J.,
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,
Young, G., Zainoun, J., Zimmer, A. and Zody, M.

Direct Submission
Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On May 12, 2000 this sequence version replaced gi:7574851.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence.submissions@genome.wi.mit.edu

----- Project Information

Center project name: L6891

Center clone name: 664_G18

Db 50659 CCGGTCTTGAGTCTTCCACAGAGTCCAGCAGCGGTGTCAGCGGTACTCTGTTGAG 50600
Qy 301 AACAGCGATCAGCCCATGACTGACCAAGTCCAAAGCAAGAAAGCCGAGTCTCTCGCA 360
Db 50599 AACAGCGATCAGCCCATGACTGACCAAGTCCAAAGCAAGAAAGCCGAGTCTCTCGCA 50540
Qy 361 GCACAATCTTGATGTCCTCCACCTCAGAGCAGCGTCTGTCTGACATCCGACATGTC 420
Db 50539 GCACAATCTTGATGTCCTCCACCTCAGAGCAGCGTCTGTCTGACATCCGACATGTC 50480
Qy 421 AAAGTCTCTCCCAAGCAGCAGCCCAAGCAGCGTCTCTCTCCAGCGTCTCCCGCCATG 480
Db 50479 AAAGTCTCTCCCAAGCAGCAGCCCAAGCAGCGTCTCTCTCCAGCGTCTCCCGCCATG 50420
Qy 481 AAGCTGGAATGATGTCAGCGCGCTTTGAGGATGTCCTCAGTGAAGTCTCAACTTTGCAT 540
Db 50419 AAGCTGGAATGATGTCAGCGCGCTTTGAGGATGTCCTCAGTGAAGTCTCAACTTTGCAT 50360
Qy 541 AAAGAAAGGCGCGAGTCCAACTGGAATCTCAGCATCTCTGATCTCAAGCCAG 600
Db 50359 AAAGAAAGGCGCGAGTCCAACTGGAATCTCAGCATCTCTGATCTCAAGCCAG 50300
Qy 601 TTTGCTCGAGCTCTCCAGACATCAGAGGCAAAATACCTGCTCTGATCTGGGCCA 660
Db 50299 TTTGCTCGAGCTCTCCAGACATCAGAGGCAAAATACCTGCTCTGATCTGGGCCA 50240
Qy 661 CAAGAGCGTATGCAAACTCTTAAGTTTACGGGACTCTCAATGACCACTATCAGTCACTGG 720
Db 50239 CAAGAGCGTATGCAAACTCTTAAGTTTACGGGACTCTCAATGACCACTATCAGTCACTGG 50180
Qy 721 CTGGCAACGTCAGTACAGCTTAAAGTAAAGCGGGGCAAAATTTCTGAAAAACATG 780
Db 50179 CTGGCAACGTCAGTACAGCTTAAAGTAAAGCGGGGCAAAATTTCTGAAAAACATG 50120
Qy 781 GACAAAGCCACCCCATCTTTTATGAGTACTGCTGCTCCAGTTCAGAACCCCTTCT 840
Db 50119 GACAAAGCCACCCCATCTTTTATGAGTACTGCTGCTCCAGTTCAGAACCCCTTCT 50060
Qy 841 ACCTACATCAGTCACTAGAAATCTCACTGGGTTTCCAAATGAAGGACATGACCGCTTG 900
Db 50059 ACCTACATCAGTCACTAGAAATCTCACTGGGTTTCCAAATGAAGGACATGACCGCTTG 50000
Qy 901 TCAGTGGACGCAAGCAAGTGGAGCAGAGATCTCCCGGTATCTCGGCTCAGAGG 960
Db 49999 TCAGTGGACGCAAGCAAGTGGAGCAGAGATCTCCCGGTATCTCGGCTCAGAGG 49940
Qy 961 TCTCAGAAACAATAGTTCGCGAGAGGACAGACTCTAAATCAAGTGAAGTTGTC 1020
Db 49939 TCTCAGAAACAATAGTTCGCGAGAGGACAGACTCTAAATCAAGTGAAGTTGTC 49880
Qy 1021 TGTCCGACATTTGTGAGCAAAATCGCGTAAACCTCCACCTAAAGCAAAACGACAGCAAG 1080
Db 49879 TGTCCGACATTTGTGAGCAAAATCGCGTAAACCTCCACCTAAAGCAAAACGACAGCAAG 49820
Qy 1081 TCACCCGAACACCATTCAGTTTGTATACAGACGTTGATGAAGATAGCTCTGCAGG 1137
Db 49819 TCACCCGAACACCATTCAGTTTGTATACAGACGTTGATGAAGATAGCTCTGCAGG 49763

RESULT 8
AC004505
LOCUS
DEFINITION Homo sapiens chromosome 20, p1 clone 86C1 (LBNL H65), complete sequence.
AC004505
VERSION AC004505.1 GI:2996633
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 57493)

AUTHORS Gray, J.W., Collins, C., Kimmerly, W., Bondoc, M., Cheng, J., Connolly, K.S., Gunning, K.M., Davis, C.A., Kadner, K., Miguel, T., Pitluck, S., Pollard, M., Rojeski, H., Subramanian, S. and Martin, C.H.
TITLE Sequencing of human chromosome 20
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 57493)
Ricke, D.O.
TITLE Large Scale Sequence Analysis and Annotation with the Sequence Comparison Analysis (SCAN) System
JOURNAL Unpublished
AUTHORS 3 (bases 1 to 57493)
Gray, J.W., Collins, C., Kimmerly, W., Bondoc, M., Cheng, J., Connolly, K.S., Gunning, K.M., Davis, C.A., Kadner, K., Miguel, T., Pitluck, S., Pollard, M., Rojeski, H., Subramanian, S. and Martin, C.H.
TITLE Direct Submission
JOURNAL Submitted (30-MAR-1998) Human Genome Center, DOE Joint Genome Institute, Lawrence Berkeley National Laboratory, MS 74-157, Berkeley, CA 94720, U.S.A.
REFERENCE 4 (bases 1 to 57493)
Kimmerly, W., Bondoc, M., Cheng, J., Connolly, K.S., Gunning, K.M., Davis, C.A., Kadner, K., Miguel, T., Pitluck, S., Pollard, M., Rojeski, H., Subramanian, S. and Martin, C.H.
TITLE Direct Submission
JOURNAL Submitted (30-MAR-1998) Human Genome Center, DOE Joint Genome Institute, Lawrence Berkeley National Laboratory, MS 74-157, Berkeley, CA 94720, U.S.A.
REFERENCE 5 (bases 1 to 57493)
Ricke, D.O.
TITLE Direct Submission
JOURNAL Submitted (20-JUN-1998) DOE Joint Genome Institute
COMMENT Sequence submitted by:
DOE Joint Genome Institute.
FEATURES
source
1. 57493
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="20"
/map="20q13.2"
/clones="86C1"
/note="LBNL H65"
complement(408..1147)
/rpt_family="LTR1"
repeat_region
complement(1436..1598)
/rpt_family="MIR"
1911..2091
/rpt_family="Alu"
complement(2129..2293)
/rpt_family="L1"
complement(3057..3173)
/note="GRAIL 2 excellent exon, frame 1"
4785..5093
/rpt_family="Alu"
5943..6244
/rpt_family="Alu"
complement(6587..6677)
/note="GRAIL 2 excellent exon, frame 2"
complement(7201..7264)
/note="95% identity B76563"
10223..10384
/note="GRAIL 2 excellent exon, frame 1"
11317..11682
/rpt_family="THE1B"
complement(13299..13486)
/rpt_family="WER3"
complement(13751..14053)
/rpt_family="Alu"
14257..14288
/note="(AC)16"
/rpt_type=tandem
/rpt_unit="ac"
14764..14853
/rpt_family="MIR"

QY 548 AAGGCGGAGTCCAACTGGAATCCTCAGCATCTTCTGATTTCTACAAGCCAGTTGGCT 607
 Db 2370 AAGGCGGAGTCCAACTGGAATCCTCAGCATCTTCTGATTTCTACAAGCCAGTTGGCT 2429

QY 608 CGAGCTCTTCCAGACATCAGAGGGCAATACCTGCTGCTGATCTGGGCCACAAAGAGC 667
 Db 2430 CGAGCTCTTCCAGACATCAGAGGGCAATACCTGCTGCTGATCTGGGCCACAAAGAGC 2489

QY 668 GTATGCAATCTTAAGTTTACGGGACTCTCAATGACCACTATACAGTCACTGGTGCCA 727
 Db 2490 GGATGCAATCTTAAGTTTACGGGACTCTCAATGACCACTATACAGTCACTGGTGCCA 2549

QY 728 AGTCAAGTACCAGCTTAGGAAACGGGGGACAAATTTCTGAAAACATGGACAAG 787
 Db 2550 AGTCAAGTACCAGCTTAGGAAACGGGGGACAAATTTCTGAAAACATGGACAAG 2609

QY 788 GCCACCCCATCTTTTATGTCAGTGAATCTGCTCCAGTTCAGAACCCCTTCTACTACA 847
 Db 2610 GCCACCCCATCTTTTATGTCAGTGAATCTGCTCCAGTTCAGAACCCCTTCTACTACA 2669

QY 848 TCAGTCACTTAGAATCTCACCTGGGTTTCCAAATGAAGACATGACCCGCTTGTGAGTG 907
 Db 2670 TCAGTCACTTAGAATCTCACCTGGGTTTCCAAATGAAGACATGACCCGCTTGTGAGTG 2729

QY 908 ACCACAAGCAAGCTGGAGCAAGATCTCCCGGTATCTCGGCTCAGAGGTTCCAG 967
 Db 2730 ACCACAAGCAAGCTGGAGCAAGATCTCCCGGTATCTCGGCTCAGAGGTTCCAG 2789

QY 968 AAACAATAGTCCGGAAGGACACAGACTCTTAAATTCAGATGTAAGTTGTCTCGGA 1027
 Db 2790 AAACAATAGTCCGGAAGGACACAGACTCTTAAATTCAGATGTAAGTTGTCTCGGA 2849

QY 1028 CATTGTGAGCAAAATCGCGTAAACTCCACTAAGCAAAACGACAGCAAGTCAACCG 1087
 Db 2850 CATTGTGAGCAAAATCGCGTAAACTCCACTAAGCAAAACGACAGCAAGTCAACCG 2909

QY 1088 RACACATTCAGTTTGTACAGACGTTGATGAGATAGTCTGCGAG 1137
 Db 2910 RACACATTCAGTTTGTACAGACGTTGATGAGATAGTCTGCGAG 2959

RESULT 11
 AL731822
 LOCUS Mouse DNA sequence from clone RP23-105H6 on chromosome 2, complete
 DEFINITION sequence.
 ACCESSION AL731822
 VERSION AL731822.31 GI:27368269
 KEYWORDS HTG.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 218037)
 Hopkins, B.
 Direct Submission
 Submitted (19-DEC-2002) Wellcome Trust Sanger Institute, Hinxton,
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
 humquery@sanger.ac.uk
 On Dec 23, 2002, this sequence version replaced gi:2762683.
 Sequence from the Mouse Genome Sequencing Consortium whole genome
 shotgun may have been used to confirm this sequence. Sequence data
 from the whole genome shotgun alone has only been used where it has
 a phred quality of at least 30.
 ----- Genome Center
 Center: Wellcome Trust Sanger Institute
 Center code: SC
 Web site: <http://www.sanger.ac.uk>
 Contact: humquery@sanger.ac.uk

During sequence assembly data is compared from overlapping clones.
 Where differences are found these are annotated as variations
 together with a note of the overlapping clone name. Note that the

variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
 This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.
 The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Emi, EMBL; Swi, SWISSPROT; Tr, TrEMBL; Wp, WormPEP; Information on the WormPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-105H6 is from the RP23-23 Mouse PAC Library constructed by the group of Pieter de Jong.
 For further details see <http://www.chori.org/bacpac/home.htm>

VECTOR: pBAC3.6.

FEATURES
source

Location/Qualifiers
 1. 218037
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10090"
 /chromosome="2"
 /clone="RP23-105H6"
 /clone_lib="RPC1-23"

ORIGIN

Query Match 70.6%; Score 909.2; DB 10; Length 218037;
 Best Local Similarity 88.4%; Pred. No. 2.6e-245;
 Matches 999; Conservative 0; Mismatches 128; Indels 3; Gaps 1;

QY 11 AGAAGGAGAAACCCAGCCCTGGAGCCACATCTGCTCTGAGCAATGGGTGCGCCCTCG 70
 Db 13029 AGAAGGAGAAACCCAGCCCTGGAGCCCTGGAGCCACATCTGCTCTGAGCAATGGGTGCGCCCTCG 13088

QY 71 CCAACCAAGCCCGCCGCTGCCATGCAATCAACCACTCAGCGCCCTGCGTCCGTCCTGA 130
 Db 13089 CTAACCAACCCCGCTGCCATGCAATCAACCACTCAGCGCCCTGCGTCCGTCCTGA 13148

QY 131 ACATCACTTGGCAAGCCAGGAGCCCTTGGCTCACTTCTGCTCCAGCCCAAGTT 190
 Db 13149 ACAACCACTTGGCAAGCCAGGAGCCCTTGGCTCACTTCTGCTCCAGCCCAAGTT 13208

QY 191 CAAGCACAATTTTCCATGTTCCACAAAGTCGAATCTCAATGTCATGACAAAGCGGTCCTGA 250
 Db 13209 CAAGCACAATTTTCCATGTTCCACAAAGTCGAATCTCAATGTCATGACAAAGCGGTCCTGA 13268

QY 251 GTCCTGCTCCACAAG---GTCAGCCAGCGTGTCCAGCGCTACCTGTTTGAGAACAGCG 307
 Db 13269 GTCCACCTCCACAGGCGCGCGCCAGCGTGGCCAGACACTACCTGTTTGAGAACAGCG 13328

QY 308 ATGAGCCCATTTGACTGACCAAGTCCAAAGCAAGAGCCAGTCTCTCGCAAGCAAT 367
 Db 13329 ACCAGCCCATTTGACTGACCAAGTCCAAAGCAAGAGCCAGTCTCTCGCAAGCAAT 13388

QY 368 CTGTGATGTCCCACTCCAGAACGACGCTCTGTCTGACATCCCGACATGTTCAAGTCC 427
 Db 13389 CTGTGATGTCCCACTCCAGAACGACGCTCTGTCTGACATCCCGACATGTTCAAGTCC 13448

QY 428 TCCCAAGCCACCAACCCCAAGCCAGCTCTCTCTCCAGGGTCCCCCCCATGAAGCTGG 487
 Db 13449 TCCCAAGCCACCAACCCCAAGCCAGCTCTCTCTCCAGGGTCCCCCCCATGAAGCTGG 13508

QY 488 AAATGAGTGTGAGCGCTTTGAGGATGTCTCAGTGAAGTCTCAACTTTGCAATAAAGAA 547
 Db 13509 AAATGAGTGTGAGCGCTTTGAGGATGTCTCAGTGAAGTCTCAACTTTGCAATAAAGAA 13568

QY 548 AAGCGCGGAGTCCAACTGGGATCTCAGCATCTTCTGATTTCTACAAGCCAGTTGGCT 607
 Db 13569 AAGCGCGGAGTCCAACTGGGATCTCAGCATCTTCTGATTTCTACAAGCCAGTTGGCT 13628

QY 608 CGAGGCTTCTCAGACATCAGAGGGCAATACCTGCTGCTGATCTGGGCCCCACAGAGC 667
Db 13629 CGAGGCTTCTCAGACCTCAGAGGGCAATACCTGCTGCTGATCTGGGCCCCACAGAGC 13688
QY 668 GTATGCAATCTCTAAGTTTACGGGACTCTCAATGACCACTATCAGTCTACCTGGTGGCA 727
Db 13689 GGATGCAATCTCTAAGTTTACGGGACTCTCAATGACCACTATCAGTCTACCTGGTGGTA 13748
QY 728 AGTCAAGTACAGCTTAGGAAACGGGGGACAAATTTCTGAAAACATGGGCAAG 787
Db 13749 AGTCAATACAGCTTAGGAAACAGGTTGGGACAAAGTTCTGAAAACATGGGCAAG 13808
QY 788 GCACCCCATCTTTATTTCAGTGACTGCTGCCCTCCAGTTTCAGAACCCCTTCTACTACA 847
Db 13809 GCACCCCATCTTTATTTCAGTGACTGCTGCCCTCCAGTTTCAGAACCCCTTCTACTACA 13868
QY 848 TCAGTCACTTAGAATCTACCTGGGTTTCCAAATGAAGGACATGACCCCTTGTTCAGTGG 907
Db 13869 TCAGGCACTTAGAGTCTACCTGGGTTTCCAAATGAAGGACATGACCCCTTGTTCAGTGG 13928
QY 908 ACCAGCAAGCAAGTGGAGCAGAGATCTCCGGGTATCTCGGCTCAGAGTCTCCAG 967
Db 13929 ACCAGCAAGCAAGTGGAGCAGAGATCTCCGGGTATCTCGGCTCAGAGTCTCCAG 13988
QY 968 AAACAATAGCTGCGGAGGACACAGACTCTAAATTTCAAGTGTAGTGTCTGTCGGA 1027
Db 13989 AAACAATAGCTGCGGAGGACACAGACTCTAAATTTCAAGTGTAGTGTCTGTCGGA 14048
QY 1028 CATTTGTGAGCAACATGCGGTAATAAATCCACTAAGCAAAACGACAGCAAGTCAACCG 1087
Db 14049 CATTTGTGAGCAACATGCGGTAATAAATCCACTAAGCAAAACGACAGCAAGTCAACCG 14108
QY 1088 AACACATTCACAGTTTGAACAGAGCTGGATGAGAAATAGTCTGCAGG 1137
Db 14109 AACACATTCACAGTTTGAACAGAGCTGGATGAGAAATAGTCTGCAGG 14158

RESULT 12

BC079877 4041 bp mRNA linear ROD 16-AUG-2004
LOCUS Mus musculus cDNA clone MGC:105242 IMAGE:6850908, complete cds.
DEFINITION BC079877
ACCESSION BC079877.1 GI:51260242
VERSION
KEYWORDS MGC.
SOURCE Mus musculus
ORGANISM Mus musculus (house mouse)

REFERENCE 1 (bases 1 to 4041)
AUTHORS Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
Klauser R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L.,
Scheetz T.E., Brownstein M.J., Ustin T.B., Toshiyuki S.,
Carninci P., Prange C., Raja S., Loquellano N.A., Peters G.J.,
Abramson R.D., Mullany S.J., Bosak S.A., McEwan P.J.,
McKernan K.J., Malek J.A., Gunaratne P.H., Richards S.,
Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S.,
Sanchez A., Whiting M., Madan A., Young A.C., Shevchenko Y.,
Bouffard G.G., Blakesley R.W., Touchman J.W., Green E.D.,
Dickson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
Butterfield Y.S., Krzywinski M.I., Skalska U., Smallos D.E.,
Schnerch A., Schein J.E., Jones S.J. and Marra M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
12477932
REFERENCE 2 (bases 1 to 4041)

AUTHORS
TITLE
JOURNAL

Director MGC Project.
Direct Submission
Submitted (09-AUG-2004) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA

REMARK
COMMENT

NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgapbs-remail.nih.gov
Tissue Procurement: Dr. Jim Lin, University of Iowa
CDNA Library Preparation: M. Bento Soares, University of Iowa
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: University of Iowa, Dr. M. Bento Soares and Dr.
Thomas L. Casavant.
Web site: <http://genome.uiowa.edu>
Contact: bento-soares@uiowa.edu; tom-casavant@uiowa.edu
Bonalder, M.F., Akabogu, I., Bair, T., Bair, J., Crouch, K., Davis, A.,
Fiehler, K., Keppel, C., Kucaba, T., Lebeck, M., Melo, A., Schaefer, K.,
Scheetz, T., Smith, C., Snit, E., Tack, D., Trout, K., Walters, J.,
Casavant, T., Soares, M.B.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: Place: Row: Column: 0
This clone was selected for full length sequencing because it
passed the following selection criteria: Hexamer frequency ORF
analysis.

FEATURES
source

Location/Qualifiers
1..4041
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="MGC:105242 IMAGE:6850908"
/tissue_type="Brain, mouse, 13.5,14.5,16.5,17.5 dpc"
/clone_lib="NIH_BMAP_FY0"
/lab_host="DH10B"
/note="Vector: pYX-ASC"
438..3530
/codon_start=1
/product="unknown (protein for MGC:105242)"
/protein_id="AAH79877.1"
/db_xref="GI:51260243"
/translation="MPRRKQQAQKRAAGYAQAEVLKEEEIEKEEEEDSGSVAHQH
SSNDGTDELETPGEQYFSCNSPGSHLSNQDAENSLGSDVSDVSKVCGR
DVSDAKTHPKLPSEPHNCMDKMTAVANILSDYSWGLGFLKLSNERNKCTDR
SGKNDPWHQDALSKSLQNLPSRSVSKPSLFSVQLYRQSKLGSVFTGASRRC
RQCSAAYDTLVLTVMNETGHYQDNRKRLRPTSYSKMVTAKKRVFDVNRPCSP
KCMFCGDFSLQDLSVHMIKTKHYQKVPKEPVTISSKMTAKKRVFDVNRPCSP
DSTTGLADSFSSQKSNLQPSNRYGYNGASYTWQFEACKSLKMECGSGSDT
LQOLTMMVTGHFLKVTSSASKGLQVLDPVLAVERQSLSETPNSLAPKPSNS
PSECTASTTEKESKEGEGEDQGVKSEDESLQKLPDPTIKYOLREEDLED
GSKGGDILKSLNVTVAINKAONGASWSAYPSIHAAYQLSECTKPMGMSOILQ
IRPNLANKRLPAPKMGMPGPVPTSLALITYQVKETEDKDEVVQCKESPHBEAT
SPSQEPGFSKLEPPSRKABPCPLKEEKPQKEPPEPVSLSLHVDPKVSPT
ALPSINPLSALQVNLHLGKATELRSPSSNSSTSPVFKSLSHVDPKVSPT
STRPAASVARYHLPENTQPIDLTQSKRAESSQAQCTSPQKHALCDIADMKVVL
PKATTPKPAASRVPMKLEIDVRFEDVSSTLHKRQSNWPNHLLILOAQF
ASSLQTSSEKYLSDLOPQERMQISKEGTSMTTISHLANVYQLKRTGKTKLN
MDKHPIFYCSDCAQKFPPTSYIISHLSHLGFGMKDMTRMAADQSKVEQISRVSS
AQRSPETIAGEEDTDSKFKCLCRRTFTVSKHAAVLHLHLSKTHSKSPHHHSQFVADVEE
"

CDS

ORIGIN

Query Match 70.5%; Score 908.2; DB 10; Length 4041;
Best Local Similarity 88.4%; Pred. No. 3.1e-245;
Matches 998; Conservative 0; Mismatches 128; Indels 3; Gaps 1;
QY 11 AGAAGAGAGAAACCCAGCCCTGGAGCCACACATCTGCTCTGAGCAATGGTGGCCCTCG 70
Db 2410 AGAAGAGAGAAACCGGAGCCCTTAGAACACAGTATCTTCTCTGACCAATGGTGGCCCG 2469
QY 71 CCAACACGCGCCCGCCCTTGCCATGCATCAACCCACTCAGCCCTGCAGTCCGTCCTGA 130

```
Db 2470 CTRACACACCCCGCTGCTCCCTTCATCAACCCCGCTAGTGCCTCTGAGTCCGTCCTGTA 2529
Qy 131 ACAATCACTTGGCAAAAGCCAGGACCCCTTGGCTTCACTTCTGCTCCAGCCCAAGTT 190
Db 2530 ACAACCACTGGGCAAAAGCCAGGACCCCTTGGCTTCACTTCTGCTCCAGTCCCAACT 2589
Qy 191 CAAAGCACATTTCCATGTTCCAAAGTCCAAATCTCAATGTCTATGCAACAGCGGCTCTTGA 250
Db 2590 CAAAGCACAGCCCGTGTTCATTAATCAAGCTCCAGTGTGACAAAGCGGTCATAA 2649
Qy 251 GTCCTGCTCCACAAG---GTCAAGCAGCGTGTCCAGGCGTACCTGTTGAGAACAGCG 307
Db 2650 GTCCCACTCCACAGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2709
Qy 308 ATCAGCCCATGACCTGACCAAGTCCAAAGCAAGAAAGCGAGTCTTCGCAAGCACAAAT 367
Db 2710 ACCAGCCCATGACCTGACCAAGTCCAAAGCAAGAAAGCGAGTCTTCGCAAGCACAAAT 2769
Qy 368 CTTGTATGTCCTCCACTCAGAACGACGCTCTGTCTGACATCCCGACATGTCAAAGTCC 427
Db 2770 CCTGACGCTCCCACTCAGAACGACGCTCTGTGTGACATTCGTATGTCAGAGTCC 2829
Qy 428 TCCCAAAAGCCACACCCCAAGCCAGCTCCTCTCCAGGTCCTCCCAAGTCCCAAGCTGG 487
Db 2830 TCCCAAAAGCCACACCCCAAGCCAGCTCCTCTCCAGGTCCTCTCCAGGTCCTCTATGAAGCTGG 2889
Qy 488 AAATGATGTCTAGGCGCTTTGAGGATGTCTCCAGTGAAGTCTCAACTTTTGCATAAAGAA 547
Db 2890 AAATGACGTCTAGCGCTTCGAGGATGTTTCGAGGAAAGTCTCCACTTTGACAAAGAA 2949
Qy 548 AAGGCGGAGTCCAACTGGAATCTCTAGATCTCTTCTGATCTTCAAGCCAGTTTGGCT 607
Db 2950 AAGGCGGAGTCCAACTGGAATCTCTAGATCTCTTCTGATCTTCAAGCTCAGTTTGGCT 3009
Qy 608 CGAGCTCTTCCAGACATCAGAGGCAATACCTGCTGTCTGATCTGCGGCCCAAGAGC 667
Db 3010 CGAGCTCTTCCAGACATCAGAGGCAATACCTGCTGTCTGATCTGCGGCCCAAGAGC 3069
Qy 668 GTATCAAAATCTTAAGTTTACGGGACTCTCAATGACCACTATCAGTCACTGGCTGGCCA 727
Db 3070 GGATCAAAATCTTAAGTTTACGGGACTCTCAATGACCACTATCAGTCACTGGCTGGCTA 3129
Qy 728 AGTCAAGTACAGCTTAGGAAAAAGCGGCGGACAAAATTTCTGAAAAACATGGAACAAG 787
Db 3130 AGTCAAAATACAGCTTAGGAAAAAGCGGCGGACAAAATTTCTGAAAAACATGGAACAAG 3189
Qy 788 GCCACCCCATCTTTTATTCAGTGTACTGCTCCAGTTCAGAACCCCTTCTACTTACA 847
Db 3190 GGCACCCCATCTTTTATTCAGTGTACTGCTCCAGTTCAGAACCCCTTCTACTTACA 3249
Qy 848 TCAGTCACTTAGAATCTCACTGGGTTTCCAAATGAAGACATATGACCCGCTTGTCAAGTGG 907
Db 3250 TCAGTCACTTAGAATCTCACTGGGTTTCCAAATGAAGACATATGACCCGCTTGTCAAGTGG 3309
Qy 908 ACCAGCAAGCAAGTGGAGCAAGAGATCTCCCGGATATCTCGGCTCAGAGGCTCCAG 967
Db 3310 ACCAGCAAGCAAGTGGAGCAAGAGATCTCCCGGATATCTCGGCTCAGAGGCTCCAG 3369
Qy 968 AAACAATAGTCCGAAGAGGACACAGACTCTAAATTCAGTGTAAAGTGTGCTGTCCGA 1027
Db 3370 AAACAATAGTCCGAAGAGGACACAGACTCTAAATTCAGTGTAAAGTGTGCTGTCCGA 3429
Qy 1028 CATTGTGAGCAAAACATGCGGTAAACTCCACTTAAGCAAAACGACAGCAAGTCAACCCG 1087
Db 3430 CATTGTGAGCAAAACATGCGGTAAACTCCACTTAAGCAAAACGACAGCAAGTCAACCCG 3489
Qy 1088 AACACCATTCAGTTTGTAAACAGACGCTGGATGAAGAAATAGCTCTGCAG 1136
Db 3490 AACACCATTCAGTTTGTAAACAGACGCTGGATGAAGAAATAGTTTCTGCAG 3538
```

RESULT 13

```
AF207880 3294 bp DNA linear ROD 17-APR-2000
LOCUS Mus musculus teashirt 2 (Tsh2) gene, partial cds.
ACCESSION AF207880
VERSION AF207880.1 GI:7578594
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 3294)
AUTHORS Caubit,X., Core,N., Boned,A., Kerridge,S., Djabali,M. and Pasano,L.
TITLE Vertebrate orthologues of the Drosophila region-specific patterning
gene teashirt
JOURNAL Mech. Dev. 91 (1-2), 445-448 (2000)
MEDLINE 20171079
PUBMED 10704881
REFERENCE 2 (bases 1 to 3294)
AUTHORS Caubit,X., Core,N., Boned,A., Djabali,M., Kerridge,S. and Pasano,L.
TITLE Direct Submission
JOURNAL Submitted (23-NOV-1999) LGPD, CNRS Case907, Parc Scientifique de
Luminy, Marseille cedex9 13288, France
FEATURES
source
1..3294
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
<1..>3294
/gene="Tsh2"
<1..>3294
/gene="Tsh2"
/note="zinc finger protein; tsh2; related to drosophila
region specific patterning gene teashirt; contains 3
teashirt-like zinc finger domains (Cx2Cx12Hmx4H) and 2
other zinc finger domains (Cx2Cx12Hx3-4H)"
/codon_start=1
/product="teashirt 2"
/db_protein_id="AAF64095.1"
/db_xref="GI:7578595"
/misc_feature
2710..2931
/gene="Tsh2"
/note="putative; Region: homeodomain"
```

ORIGIN

```
Query Match 69.7%; Score 897.2; DB 10; Length 3294;
Best Local Similarity 98.3%; Pred. No. 3.9e-242;
Matches 987; Conservative 0; Mismatches 128; Indels 3; Gaps 1;

Qy 11 AGAAGAGAGAAACCCGAGCCCTCGGAGCCACATCTGCTCTCGAGCAATGGGTCGCGCCCTCG 70
Db 2177 AGAAGAGAGAAACCCGAGCCCTTAGAACCATGATCTCTCTGACCAATGGCTGTGCGCGG 2236
Qy 71 CCAACCAAGCCCGCCCTCGCATGTCATCAACCCACTCAGCGCCCTGCGTCCGTCCTGA 130
```

Db 2237 CTAACACACCCCGGCTGCTTCCATCAACCCGCTCAGTGCCTCGAGTCCGTCCTGA 2296
Qy 131 ACAATCACTTGGCAAGCCAGGAGCCCTTGCGCTCACCTTCTGCTCCAGCCCAAGTT 190
Db 2297 ACACACCTGGCAAGCCAGGAGCCCTTGCGCTCCCTTCTGCTCCAGTCCCAACT 2356
Qy 191 CAAGCACAATTTCCATGTTCCAAAGTCCGAATCTCAATGTCTATGTCATGGAACAAGCCGCTCTTGA 250
Db 2357 CAAGCAACAGCCCGTGTTCATAAATCCAGCTCCACGCTGGTGGACAGCCGTCATAA 2416
Qy 251 GTCCTGCTCCCAAG---GTCAGCCAGCGGTGCCAGGCGCTACCTGTTTGGAGAACAGCG 307
Db 2417 GTCCCACTCCACAGCGCGCGGCGAGGCTGGCCAGACACTACCTGTTTGGAGAACAGCG 2476
Qy 308 ATCAGCCCATTTGACCTGACCAAGTCCAAAGCAAGCAAGCCGAGTCCCTGCCAGCACAAT 367
Db 2477 ACCAGCCCATTTGACCTGACCAAGTCCAAAGCAAGCAAGCCGAGTCCCTGCCAGCACAAT 2536
Qy 368 CTTGTATGTCCTCCACCTCAGAACGACGCTCTGTCTGACATCGCCGACATGGTCAAAAGTCC 427
Db 2537 CTTGACAGTCCCACTCAGNAGCATGCTCTGTGTGACATTTGCTGATATGGTCAAGTCC 2596
Qy 428 TCCCAAAAGCCACCAACCCAAAGCCAGCCTCTCTCCAGGCTCCCGCCCATGAAGCTGG 487
Db 2597 TCCCAAAAGCCACCAACCCAAAGCCAGCCTCTCTCGAGGGTCCCTCTATGAAGCTGG 2656
Qy 488 AAATGATGTCAGGCGCTTGGAGTGTCTCCAGTGAAGTCTCAACTTTGTCATAAAGAA 547
Db 2657 AAATGACGTGACAGCGCTTCGAGGATGTTTCGAGCGAAGTCTCCATTTGTCACAAAAGGA 2716
Qy 548 AAGCCCGGAGTCAAACTGGAATCTCTCAGCATCTTCTGATTTTACAAAGCCAGTTTGCCT 607
Db 2717 AAGCCCGGAGTCAAACTGGAATCTCTCAGCATCTTCTGATTTTACAAAGCCAGTTTGCCT 2776
Qy 608 CGAGCCTCTTCCAGCAATCAGAGGGCAAAATACCTGCTGTCTGATTTGGGCCCCCAAGAGC 667
Db 2777 CGAGCCTCTTCCAGCAATCAGAGGGCAAAATACCTGCTGTCTGATTTGGGCCCCCAAGAGC 2836
Qy 668 GTATGCAATCTTAAGTTTACGGGACTCTCAATGACCACTATCAGTCTACTGGTGGCCA 727
Db 2837 GGATGCAATCTTAAGTTTACGGGACTCTCAATGACCACTATCAGTCTACTGGTGGCTA 2896
Qy 728 ACCTCAAGTACCACTTAGGAAAACGGGGGGAACAAAATTTCTGAAAAACATGGACAAAG 787
Db 2897 ACCTCAATACCAAGCTTAGGAAAACAGGTGGGACAAAGTTCTGAAAAACATGGACAAAG 2956
Qy 788 GCACCCCATCTTTTATTTGAGTGAAGTGTGCTTCCAGTTCAGAACCCCTTTTACCTACA 847
Db 2957 GGACCCCATCTTTTATTTGAGTGAAGTGTGCTTCCAGTTCAGAACCCCTTTTACCTACA 3016
Qy 848 TCAGTCACTTAGATCTCACTGGGTTTCCAAATGAAGGACATGACCCGCTGTCTAGTGG 907
Db 3017 TCAGCACCCTTAGATCTCACTGGGTTTCCAAATGAAGGACATGACCCGATGGCGGCTG 3076
Qy 908 ACCAGCAAGCAAGTGGAGCAAGAGATCTCCCGGCTATCGTCCGCTCAGAGTCTCCAG 967
Db 3077 ACCAGCAAGCAAGTGGAGCAGGAGATCTCCCGAGTGTCTGCTCGGCTCAGAGTCTCCAG 3136
Qy 968 AAACAATAGTCTCGAAGGAGCACAGACTCTAAATTTCAAGTGTAAAGTTGTGCTGTCGA 1027
Db 3137 AAACAATAGTCTCGAAGGAGCACAGACTCTAAATTTCAAGTGTAAAGTTGTGCTGTCGA 3196
Qy 1028 CATTTGTGAGCAACATCGGTTAAACTCCACCTTAAGCAAAAGCCAGCAAGTCAACCG 1087
Db 3197 CATTTGTGAGCAACATCGGTTAAACTCCACCTTAAGCAAAAGCCAGCAAGTCAACCG 3256
Qy 1088 AACACCAATTACAGTTTGTAAACAGAGCTGGATGAAGAA 1125
Db 3257 AACACCACTCTAGTTTGTAGCAGAGCTGGATGAAGAA 3294

CQ078477
LOCUS CQ078477 2012 bp DNA linear PAT 20-JAN-2004
DEFINITION Sequence 14277 from Patent WO0157278.
ACCESSION CQ078477
VERSION CQ078477.1 GI:41048346
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Penn, S.G., Hanzel, D.K., Chen, W. and Rank, D.R.
TITLE Human genome-derived single exon nucleic acid probes useful for
analysis of gene expression in human hela cells or other human
cervical epithelial cells
JOURNAL Patent: WO 0157278-A 14277 09-AUG-2001;
FEATURES
Location/Qualifiers
1..2012
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
/note="MAP TO AC004505.1-EXPRESSED IN HELA, SIGNAL = 1-NT
HIT: AF207880.1, EVALUUE 0.00e+00-EST_HUMAN HIT: H12950.1,
EVALUUE 0.00e+00-SWISSPROT HIT: P22265, EVALUUE 1.00e-07"
ORIGIN
Query Match 51.0%; Score 656.4; DB 6; Length 2012;
Best Local Similarity 99.8%; Pred. No. 4,7e-174;
Matches 657; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 GAGGGCAGCGAGAGAGAGAAACCCAGCCCTGGAGCCACACATCTGCTCTGAGCAATGGG 60
Db 1347 GAGGGCAGCGAGAGAGAGAAACCCAGCCCTGGAGCCACACATCTGCTCTGAGCAATGGG 1406
Qy 61 TCGGCTCTCGCAACCAACGCGCCCTTGCATGCATCAACCCACTCAGCGCCTGCGAG 120
Db 1407 TCGGCTCTCGCAACCAACGCGCCCTTGCATGCATCAACCCACTCAGCGCCTGCGAG 1466
Qy 121 TCGGCTCTGAAACAATCACTTGGGCAAGCCAGGAGCCCTTGGCTCAGCTCTCTGCTCC 180
Db 1467 TCGGCTCTGAAACAATCACTTGGGCAAGCCAGGAGCCCTTGGCTCAGCTCTCTGCTCC 1526
Qy 181 AGCCCAAGTTCAAGCACAAATTTCCATGTTCCACAAGTCCAAATCTCAATGTCTATGGACAAG 240
Db 1527 AGCCCAAGTTCAAGCACAAATTTCCATGTTCCACAAGTCCAAATCTCAATGTCTATGGACAAG 1586
Qy 241 CCGGCTTTGAGTCTGCTCTCCACAAGGTTCAGCGGTTCAGCGGCTTCCAGGCGCTACCTGTTTGA 300
Db 1587 CCGGCTTTGAGTCTGCTCTCCACAAGGTTCAGCGGTTCAGCGGCTTCCAGGCGCTACCTGTTTGA 1646
Qy 301 AACAGGGATCAGCCCATTTGACCTGACCAAGTCCAAAAGCAAGAAAGCCAGTCTCTCGCAA 360
Db 1647 AACAGGGATCAGCCCATTTGACCTGACCAAGTCCAAAAGCAAGAAAGCCAGTCTCTCGCAA 1706
Qy 361 GCACAATCTTGATGTCCCCACTCAGAAGCAGCTCTGCTGACATCCCGACATGGTGC 420
Db 1707 GCACAATCTTGATGTCCCCACTCAGAAGCAGCTCTGCTGACATCCCGACATGGTGC 1766
Qy 421 AAAGTCTCTCCCAAGGCCACCCCAAGCCAGCCTCTCTCTCCAGGCTCCCCCCCATTG 480
Db 1767 AAAGTCTCTCCCAAGGCCACCCCAAGCCAGCCTCTCTCTCCAGGCTCCCCCCCATTG 1826
Qy 481 AAGCTGGAAATGGATGTGAGGCGCTTTGAGGATGCTCCAGTGAAGTCTCAACTTTGCAT 540
Db 1827 AAGCTGGAAATGGATGTGAGGCGCTTTGAGGATGCTCCAGTGAAGTCTCAACTTTGCAT 1886
Qy 541 AAAAGAAAAGGCGGCGAGTCCCAACTCTCAGCATCTTCTGATTTCTACAAGCCCGAG 600
Db 1887 AAAAGAAAAGGCGGCGAGTCCCAACTCTCAGCATCTTCTGATTTCTACAAGCCCGAG 1946
Qy 601 TTTGCTCGAGCCTCTTCCAGACATCAGAGGGCAAAATACCTGCTGCTCATCTGGGCC 658
|||||

Db 1947 TTGCTCGAGCCTCTTCCAGACATCAGAGGCGCAATACCTGCTGCTGATCTGGCCC 2004

RESULT 15
BV014871 671 bp DNA linear STS 30-MAY-2003
LOCUS S212P60463FH5.TO CZECHII/Ei Mus musculus STS genomic, sequence
DEFINITION tagged site.

ACCESSION BV014871
VERSION BV014871.1 GI:31098766
KEYWORDS STS.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 671)
AUTHORS Wade, C.M., Kulbokas, E.J. III, Kirby, A.W., Zody, M.C., Mullikin, J.C., Lander, E.S., Lindblad-Toh, K. and Daly, M.J.
TITLE The mosaic structure of variation in the laboratory mouse genome
JOURNAL Nature 420 (6915), 574-578 (2002)
MEDLINE 22354684
PUBMED 12466852

COMMENT
Contact: Kerstin Lindblad-Toh
Whitehead Institute for Biomedical Research, Center for Genome
Research
320 Charles Street, Cambridge, MA 02141, USA
Tel: 6172521477
Fax: 6172580903
Email: kersti@genome.wi.mit.edu
Primer A: None
Primer B: None
STS size: 671
Protocol:
WGS-discovery: Paired-end low-coverage whole genome shotgun reads
were generated from 12981/SvImJ, C3H/HeJ, and BALB/cByJ. The WGS
reads were placed uniquely on the MGSv3 C57BL/6J assembly and SNP
detection was carried out by SSAHA-SNP. 225,000 reads were
annotated
as STSs and 81,000 SNPs were annotated with alleles from C57BL/6J
and the strain from which the particular read came. The validation
rate for these SNPs was estimated at approximately 98%.

FEATURES
source
1..671
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="CZECHII/Ei"
/db_xref="taxon:10090"
/map="4 2 22-621 170891742-170892341"
/clone_lib="CZECHII/Ei"
<1..>671

STS
ORIGIN

Query Match 44.1%; Score 568.6; DB 11; Length 671;
Best Local Similarity 90.5%; Pred. No. 2.8e-149;
Matches 607; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

Qy 363 ACAATCTTGATGTCCTCCACCTCACAAGCAGCCTCTGTCGACATCGCGGACATGGTCAA 422
Db |||||
1 ACAATCTGACGTCCTCCACCTCAGAAGCAGTCTGTCGACATGCGCGATGCGTCAA 60
|||

Qy 423 AGTCCTCCCAAGCACCACCCCAAGCCAGCCTCTCTCCAGGGTCCCCCCCATGAA 482
Db |||||
61 GGTCTCTCCCAAGCACCACCCCAAGCCAGCCTCTCTCCAGGGTCCCCCTCTATGAA 120
|||

Qy 483 GCTGAATGATGTCAGCGCTTTGAGGATGTCCTCAGTGAAGTCTCAACTTTGCATAA 542
Db |||||
121 GCTGAATTCGACGTCAGACGCTTTGAGGATGTTTCGAGCGAAGTCTCCACTTTGCACAA 180
|||

Qy 543 AAGAAAGCGCGGAGTCCAACTGGAATCCTCAGCATCTCTGATTTACAAGCCAGTT 602
Db |||||
181 AAGAAAGCGCGGAGTCCAACTGGAATCCTCAGCATCTCTCTCATCTCTGCAAGTCTAGTT 240
|||

Qy 603 TGCTCGAGCCTCTTCCAGACATCAGAGGCGCAATACCTGCTGCTGATCTGGGCCACCA 662
Db |||||
241 TGCTCGAGCCTCTTCCAGACCTCAGAGGCGCAATACCTGCTTCTGACCTGGGCCCCCA 300
|||

Qy 663 AGAGCGTATGCAAAATCTTAAGTTTACGGGACTCTCAATGACCACTATCAGTCTAGGCT 722
Db |||||
301 AGAGCGATGCAAAATCTTCAAGTTTACCGGACTCTCCATGACCAATCAGCCACTGGCT 360
|||

Qy 723 GGCCAACGTCAAGTACCAAGTTAGGAAAAAGGGGGGACAAAAATTTCTGAAAAACATGGA 782
Db |||||
361 GGCTTAACGTCAAAATACCAAGTTAGGAAAAAGGGTGGGACAAAGTTCTCTGAAAAACATGGA 420
|||

Qy 783 CAAGGGCCACCCCATCTTTTATTCAGTGTGCTGCTCCAGTTCCAGAACCCCTTCTAC 842
Db |||||
421 CAAGGGCCACCCCATCTTTTATTCAGTGTGCTGCTCCAGTTCCAGAACCCCTTCTAC 480
|||

Qy 843 CTACATCAGTCACCTAGAAATCTCACCTGGGTTTCCAAATGAAGGACATGACCCGTTGTC 902
Db |||||
481 CTACATCAGCCACCTAGAGTCTCACCTGGGTTTCCAAATGAAGGACATGACCCGATGTC 540
|||

Qy 903 AGTGGACCAAGCAAGCAAGTGGAGCAAGAGATCTCCCGGATTCGTCGGTCTCAGAGTTC 962
Db |||||
541 GGCTGACCAAGCAAGTGGAGCAAGAGATCTCCCGAGTGTCTCGGCTCAGAGTTC 600
|||

Qy 963 TCCAGAAACAATAGCTGCCGAGGACACAGACTCTAAATTTCAAGTGTAAAGTTGCTGCTG 1022
Db |||||
601 TCCAGAAACAATAGCTGCCGAGGACACAGACTCTAAATTTCAAGTGTAAAGTTGCTGCTG 660
|||

Qy 1023 TCGGACATTTG 1033
Db |||||
661 TCGGACATTTG 671
|||

Search completed: May 5, 2005, 03:58:23
Job time : 6299.78 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 4, 2005, 20:02:41 ; Search time 891.572 Seconds
(without alignments)
8551.893 Million cell updates/sec

Title: US-08-731-499-5
Perfect score: 1288
Sequence: 1 GAGGCGAGGAGAGAGAA.....AAAAAATTC 1288

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_16Dec04.*
1: Geneseqn1980s.*
2: Geneseqn1990s.*
3: Geneseqn2000s.*
4: Geneseqn2001as.*
5: Geneseqn2001bs.*
6: Geneseqn2002as.*
7: Geneseqn2002bs.*
8: Geneseqn2003as.*
9: Geneseqn2003bs.*
10: Geneseqn2003cs.*
11: Geneseqn2003ds.*
12: Geneseqn2004as.*
13: Geneseqn2004bs.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1288	100.0	1288	2 AAV04700	Homo sapi
2	1268	98.4	3313	10 ABZ79883	Human nuc
3	1265	98.2	3322	10 ADB62219	Human CDN
4	1013.6	78.7	3442	10 ADC30023	Human nov
5	773	60.0	2449	4 AAS26449	Human CDN
6	773	60.0	2449	8 ABX73790	Human nov
7	770.4	59.8	799	4 AAS26014	Human CDN
8	770.4	59.8	799	8 ABX73355	Human nov
9	656.4	51.0	2012	4 AAI24344	Human nov
10	521	40.5	1964	4 AAI15164	Probe #14
11	390.4	30.3	3604	4 AAI2638	Human bre
12	390.4	30.3	3604	4 AAI06732	Human rep
13	390.4	30.3	3907	6 ABK34640	Human CDN
14	390.4	30.3	7081	10 ADF55427	Human DNA
15	390	30.3	1079	10 ADC31952	Human nov
16	387.2	30.1	4854	13 ADR06933	Full leng
17	377.8	29.3	2918	2 AAX40074	Colon can
18	272	21.1	3692	12 ADQ20133	Human sof
19	272	21.1	4407	5 ABX71366	Human tra
20	272	21.1	5065	12 ADQ23080	Human sof

21	272	21.1	5065	12 ADQ24468	Human sof
22	272	21.1	5065	12 ADQ24179	Human sof
23	256.4	19.9	1144	2 AAV61044	Nucleotid
24	157	12.2	346	6 ABK64270	Human ben
25	143.8	11.2	1999	4 AAK85937	Human imm
26	143.4	11.1	2044	4 AAS25958	Human CDN
27	143.4	11.1	2044	8 ABX73299	Human nov
28	143	11.1	852	4 AAS26397	Human CDN
29	143	11.1	852	8 ABX73738	Human nov
30	105	8.2	3370	10 ADB62184	Human CDN
31	97.2	7.5	450	9 ACH12952	Human adu
32	65.2	5.1	537	9 ACH32461	Human end
33	63.8	5.0	1258	4 AAK85936	Human imm
34	63.2	4.9	555	4 AAK63792	Human imm
35	51.6	4.0	636	4 AAL15974	Human bre
36	51.6	4.0	640	11 ACN92451	Breast ca
37	51.6	4.0	651	4 AAL24818	Human bre
38	51.6	4.0	904	11 ACN85935	Breast ca
39	46.4	3.6	2000	8 ADA71938	Rice gene
40	41.6	3.2	258	11 ADI31447	Human CDN
41	40	3.1	515	12 ACH73311	Human gen
42	40	3.1	2000	8 ADA71938	Rice gene
43	38.6	3.0	6643	12 ADQ83730	Human tum
44	38.2	3.0	677	2 AAX39783	Gastric c
45	38.2	3.0	4590	5 AAH24065	Yeast AOD

ALIGNMENTS

RESULT 1
AAV04700
ID AAV04700 standard; cDNA; 1288 BP.
XX
AC AAV04700;
XX
DT 21-JUL-1998 (first entry)
XX
DE Homo sapiens 20q13 amplicon 41.1 transcript.

20q13 amplicon; chromosome 20; tumour; detection; homeobox T shirt;
chromosomal abnormalities; probe; gene therapy; antisense inhibition;
treatment; age-related macular degeneration; retinitis pigmentation;
Leber's congenital amaurosis; ds.

Homo sapiens.

WO9802539-A1.

22-JAN-1998.

15-JUL-1997; 97WO-US012343.

15-JUL-1996; 96US-00680395.

16-OCT-1996; 96US-00731499.

17-JAN-1997; 97US-00785532.

(REGC) UNIV CALIFORNIA.

Gray JW, Collins CC, Hwang S, Godfrey T, Kowbel D, Rommens J;

WPI; 1998-110587/10.

New sequences from the 20q13 amplicon - used for detecting chromosomal
abnormalities, particularly tumours, and for developing products for
treating diseases.

Claim 1; Page 62; 91pp; English.

The sequence is that of a cDNA sequence 41.1, which was isolated from the
20q13 amplicon. It shows homology to the homeobox T shirt gene in
Drosophila. It can be used as a probe for the detection of chromosomal
abnormalities at 20q13. It and other sequences isolated from the 20q13

CC	amplicon are consistently amplified in primary tumours. These sequences									
CC	are useful as probes or as probe targets for monitoring the relative copy									
CC	number of corresponding sequences from a biological sample such as tumour									
CC	cells. The sequences can also be used in therapeutic applications for									
CC	modulating the expression of the endogenous gene or the activity of the									
CC	gene product. Examples of therapeutic approaches include antisense									
CC	inhibition of gene expression, gene therapy, and monoclonal antibodies									
CC	that specifically bind the gene products. The products can also be used									
CC	in the treatment of other diseases, e.g. age-related macular									
CC	degeneration, Leber's congenital amaurosis and retinitis pigmentata									
XX										
SQ	Sequence 1288 BP; 355 A; 396 C; 274 G; 263 T; 0 U; 0 Other;									
	Query Match	100.0%;	Score 1288;	DB 2;	Length 1288;					
	Best Local Similarity	100.0%;	Pred. No. 0;							
	Matches 1288;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;					
QY	1	GAGGGCAGCAGAGAGAGAAACCCAGCCCTCTGAGGCCACATCTGCTCTGAGCAATGGG	60							
Db	1	GAGGGCAGCAGAGAGAGAAACCCAGCCCTCTGAGGCCACATCTGCTCTGAGCAATGGG	60							
QY	61	TGGGCCCTCGCCAAACACGCGCCCGCCCTGGCATGCATCAACCCACATCAGGCGCCCTGCAG	120							
Db	61	TGGGCCCTCGCCAAACACGCGCCCGCCCTGGCATGCATCAACCCACATCAGGCGCCCTGCAG	120							
QY	121	TCCGTCTCTGAACAAATCACTTTGGGCAAAAGCCACGGAGGCCCTTGGCGCTCACCTTCTCTGCTCC	180							
Db	121	TCCGTCTCTGAACAAATCACTTTGGGCAAAAGCCACGGAGGCCCTTGGCGCTCACCTTCTCTGCTCC	180							
QY	181	AGGCCAAGTTCAAGCACAATTTTCATGTTTCCACAAGTCGAATCTCAATGTCTATGGACAAG	240							
Db	181	AGGCCAAGTTCAAGCACAATTTTCATGTTTCCACAAGTCGAATCTCAATGTCTATGGACAAG	240							
QY	241	CCGGTCTTGAGTCCTGCTCCACAAGGTCAGCCAGCGGTGCCAGGCGGTACCTGCTTTTCAG	300							
Db	241	CCGGTCTTGAGTCCTGCTCCACAAGGTCAGCCAGCGGTGCCAGGCGGTACCTGCTTTTCAG	300							
QY	301	AACAGCGATCAGCCCATTTGACTGACCAAGTCCAAAAGCAAGAACGCGAGTCTCTCGCAA	360							
Db	301	AACAGCGATCAGCCCATTTGACTGACCAAGTCCAAAAGCAAGAACGCGAGTCTCTCGCAA	360							
QY	361	GCACAAATCTTGATGTCTCCCACTCAGAAAGCAGCTCTGTCTGATGTCGCCGACATGTC	420							
Db	361	GCACAAATCTTGATGTCTCCCACTCAGAAAGCAGCTCTGTCTGATGTCGCCGACATGTC	420							
QY	421	AAAGTCTCTCCCAAGGCCACACCCCAAGCCAGCCTCTCTCCAGGGTCCCCCCCATG	480							
Db	421	AAAGTCTCTCCCAAGGCCACACCCCAAGCCAGCCTCTCTCCAGGGTCCCCCCCATG	480							
QY	481	AAGCTGGAAATGGATGTTCAGCGGTTTGGAGATGTCTCCAGTGAAGTCTCAACTTTGCAT	540							
Db	481	AAGCTGGAAATGGATGTTCAGCGGTTTGGAGATGTCTCCAGTGAAGTCTCAACTTTGCAT	540							
QY	541	AAAAAGAAAGCCGGCAGTCCAACTGGAATCCTCAGCATCTTCTGATTTCTACAAGCCCAAG	600							
Db	541	AAAAAGAAAGCCGGCAGTCCAACTGGAATCCTCAGCATCTTCTGATTTCTACAAGCCCAAG	600							
QY	601	TTTGGCTCGAGCCTCTTCCAGACATCAGAGGGCAAAATACCTGCTGTCTGATCTGGGCCCA	660							
Db	601	TTTGGCTCGAGCCTCTTCCAGACATCAGAGGGCAAAATACCTGCTGTCTGATCTGGGCCCA	660							
QY	661	CAAGAGCGTATGCAAAATCTCTAAGTTTACGGGATCTCTCAATGACCACTATCAGTCACTGG	720							
Db	661	CAAGAGCGTATGCAAAATCTCTAAGTTTACGGGATCTCTCAATGACCACTATCAGTCACTGG	720							
QY	721	CTGGCCCAACCGTCAAGTACCACTTTAGGAAAAAGCGGGCGGACAAAATTTCTGAAAAACATG	780							
Db	721	CTGGCCCAACCGTCAAGTACCACTTTAGGAAAAAGCGGGCGGACAAAATTTCTGAAAAACATG	780							
QY	781	GACAAAGGCCACCCCATCTTTTATTTGAGTGACTGTGCTCTCCAGTTCAGAACCCCTTCT	840							
Db	781	GACAAAGGCCACCCCATCTTTTATTTGAGTGACTGTGCTCTCCAGTTCAGAACCCCTTCT	840							

PI Lal PG, Lee EA, Lee S, Lee SY, Li JX, Lu DAM, Lu Y, Lehr-Mason PM;
 PI Nguyen DB, Ramkumar J, Sprague WW, Tang YT, Thangavelu K; Yue H;
 PI Thornton M, Tran UK, Wallia NK, Warren BA, Xu Y, Yao MG, Yue H;
 PI Yue H, Zebajadian Y;
 XX
 DR WPI: 2003-256709/25.
 DR P-PSDB; ABP96228.
 XX
 PT New human nucleic acid-associated proteins polypeptide, useful for
 PT preparing a composition for diagnosing or treating e.g., cardiovascular
 PT or neurological disorders.
 XX
 PS Claim 5; Page 266-267; 290pp; English.
 XX
 CC AB279873 to AB279905 encode the human nucleic-acid associated proteins
 CC designated NAAP-1 to NAAP-33 given in ABP96250. The NAAP
 CC sequences have cardiant, cytotstatic and neuroprotective activities, and
 CC can be used in gene therapy. The NAAP sequences can be used for preparing
 CC a composition for diagnosing or treating a disease or condition
 CC associated with decreased expression or overexpression of functional NAAP
 CC e.g. cardiovascular or neurological disorders or cancer
 XX
 SQ Sequence 3313 BP; 948 A; 922 C; 812 G; 631 T; 0 U; 0 Other;

 Query Match 98.4%; Score 1268; DB 10; Length 3313;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1268; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

 QY 1 GAGGGCAGGAGGAGAGAACCCAGCCCTGGAGCCACATCTGCTCAGCAATGGG 60
 DB 2046 GAGGGCAGGAGGAGAGAACCCAGCCCTGGAGCCACATCTGCTCAGCAATGGG 2105

 QY 61 TGGGCGCTCGCAACACCGCCCGCCCTGGCATGATCAACCCACTCAGCGCCCTCGAG 120
 DB 2106 TGGGCGCTCGCAACACCGCCCGCCCTGGCATGATCAACCCACTCAGCGCCCTCGAG 2165

 QY 121 TCCGTCCTGAACAATCACTTGGGCAAGCCAGGAGCCCTTGGCTCACTTCTGCTCC 180
 DB 2166 TCCGTCCTGAACAATCACTTGGGCAAGCCAGGAGCCCTTGGCTCACTTCTGCTCC 2225

 QY 181 AGCCCAAGTTCAAGCACAATTTCCATGTTCCACAGTCAATCTCAATGTATGACCAAG 240
 DB 2226 AGCCCAAGTTCAAGCACAATTTCCATGTTCCACAGTCAATCTCAATGTATGACCAAG 2285

 QY 241 CCGGCTTTAGTCTCTGCTCCACAGGTGAGCCAGCGTGTCCAGCGCTACTGTTTGAG 300
 DB 2286 CCGGCTTTAGTCTCTGCTCCACAGGTGAGCCAGCGTGTCCAGCGCTACTGTTTGAG 2345

 QY 301 AACAGCGATCAGCCCATTTGACCTGACCAAGTCCAAAGCAAGAGCGAGTCTCTCGCA 360
 DB 2346 AACAGCGATCAGCCCATTTGACCTGACCAAGTCCAAAGCAAGAGCGAGTCTCTCGCA 2405

 QY 361 GCACATCTTGATGTCGCCACCTCAGAGCAGCTGTCTGACATCGCGCATGTC 420
 DB 2406 GCACATCTTGATGTCGCCACCTCAGAGCAGCTGTCTGACATCGCGCATGTC 2465

 QY 421 AAGGTCCTCCCAAGCCACACCCCAAGCAGCTCTCTCCAGGCTCCCCCCTATG 480
 DB 2466 AAGGTCCTCCCAAGCCACACCCCAAGCAGCTCTCTCCAGGCTCCCCCCTATG 2525

 QY 481 AAGCTGGAATGGATGTCAGGCGCTTTGAGGATGTCCTCAGTGAAGTCTCAACTTTGAT 540
 DB 2526 AAGCTGGAATGGATGTCAGGCGCTTTGAGGATGTCCTCAGTGAAGTCTCAACTTTGAT 2585

 QY 541 AAAAGAAAAGCCGCGAGTCCAACTGGAAATCTCAGCATCTTCTGATTTCAAGCCCGAG 600
 DB 2586 AAAAGAAAAGCCGCGAGTCCAACTGGAAATCTCAGCATCTTCTGATTTCAAGCCCGAG 2645

 QY 601 TTTGCTCAGGCTCTTCCAGCATCAGAGGCAATACCTGCTCTGATCTGGGCCA 660
 DB 2646 TTTGCTCAGGCTCTTCCAGCATCAGAGGCAATACCTGCTCTGATCTGGGCCA 2705

 QY 661 CAAGAGCGTATGCAAACTCTTAAGTTTACGGGAGCTCTCAATGACCACTATCAGTCACTGG 720

Db 2706 CAAGAGCGTATGCAAACTCTTAAGTTTACGGACTCTCAATGACCACTATCAGTCACTGG 2765
 QY 721 CTGCGCAACGCTCAAGTACCAAGCTTAGGAAAACGCGGCGGACAAATTTCTGAAAAACATG 780
 Db 2766 CTGCGCAACGCTCAAGTACCAAGCTTAGGAAAACGCGGCGGACAAATTTCTGAAAAACATG 2825
 QY 781 GACAAAGCCACCCCATCTTTATTGTCAGTGAAGTGTGCTCCAGTCTCAGAACCCCTTCT 840
 Db 2826 GACAAAGCCACCCCATCTTTATTGTCAGTGAAGTGTGCTCCAGTCTCAGAACCCCTTCT 2885
 QY 841 ACTTACATCAGTCACTTAGAATCTCACTGGTTTCCAAATGAAGGACATGACCCGCTTG 900
 Db 2886 ACTTACATCAGTCACTTAGAATCTCACTGGTTTCCAAATGAAGGACATGACCCGCTTG 2945
 QY 901 TCAGTGGACAGCAAAAGCAAGGTGGAGCAAGAGATCTCCCGGGTATCGTCGGCTCAGAGG 960
 Db 2946 TCAGTGGACAGCAAAAGCAAGGTGGAGCAAGAGATCTCCCGGGTATCGTCGGCTCAGAGG 3005
 QY 961 TCTCCAGAAACAATAGTGTGCGGAAGAGGACACAGACTTAAATTCAAAGTGTAAAGTTGTGC 1020
 Db 3006 TCTCCAGAAACAATAGTGTGCGGAAGAGGACACAGACTTAAATTCAAAGTGTAAAGTTGTGC 3065
 QY 1021 TGTGCGGACATTTGTGAGCAAAACATGCGGTAAACCTCCACCTTAAGCAAAACGACAGCAAG 1080
 Db 3066 TGTGCGGACATTTGTGAGCAAAACATGCGGTAAACCTCCACCTTAAGCAAAACGACAGCAAG 3125
 QY 1081 TCACCCGAAACCAATTCACAGATTTGTAAACAGACGCTGGATGAAGAAATAGCTCTGCAGGACG 1140
 Db 3126 TCACCCGAAACCAATTCACAGATTTGTAAACAGACGCTGGATGAAGAAATAGCTCTGCAGGACG 3185
 QY 1141 AATGCTTTAGTTTCCATTTCCAGCTGGATGCCCTCACACTGAACCCCTTCTGTTGCA 1200
 Db 3186 AATGCTTTAGTTTCCATTTCCAGCTGGATGCCCTCACACTGAACCCCTTCTGTTGCA 3245
 QY 1201 CCATCTGCTTCTGACATTTGAACTCAATTTGAACTCTCTGACACCCCTGGCTCTGAGAGA 1260
 Db 3246 CCATCTGCTTCTGACATTTGAACTCAATTTGAACTCTCTGACACCCCTGGCTCTGAGAGA 3305
 QY 1261 CTGCAAAA 1268
 Db 3306 CTGCAAAA 3313

 RESULT 3
 ADB62219
 ID ADB62219 standard; cDNA; 3322 BP.
 XX
 AC ADB62219;
 XX
 DT 04-DEC-2003 (first entry)
 XX
 DE Human cDNA encoding clone CTONG20074000.
 XX
 KW Human; ss; gene; pharmaceutical; diagnostic; gene therapy;
 KW tissue regeneration; cell regeneration; membrane protein;
 KW signal transduction-related protein; transcription-related protein;
 KW osteoporosis; neurological disease; cancer; tumour.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 81..3185
 FT /*tag= a
 FT /product= "Clone CTONG20074000 protein"
 XX
 PN EP1308459-A2.
 XX
 PD 07-MAY-2003.
 XX
 PF 28-MAR-2002; 2002EP-00007401.
 XX
 PR 05-NOV-2001; 2001JP-00379298.

RESULT 5
AAS26449
ID AAS26449 standard; cDNA; 2449 BP.
XX
AC AAS26449;
XX
DT 07-NOV-2001 (first entry)
XX
DE Human cDNA encoding a novel secreted protein, Seq ID 628.
XX
KW Human; immunosuppressive; antiarthritic; ss; antirheumatic; cytostatic;
KW cardiant; vasotropic; cerebroprotective; nootropic; neuroprotective;
KW antibacterial; virucide; fungicide; ophthalmologic; vulnerary;
KW secreted protein; rheumatoid arthritis; hyperproliferative disorder;
KW cardiovascular disorder; cardiac arrest; cerebrovascular disorder;
KW cerebral ischaemia; angiogenesis; nervous system disorder;
KW Alzheimer's disease; infection; ocular disorder; corneal infection;
KW wound healing; epithelial cell proliferation; skin ageing; food additive;
KW preservative; antiproliferative.
OS Homo sapiens.
XX
PN WO20015322-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US0001341.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226686P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0232403P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239335P.
PR 13-OCT-2000; 2000US-0239337P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.

PR 17-NOV-2000; 2000US-0249244P.
 PR 17-NOV-2000; 2000US-0249245P.
 PR 17-NOV-2000; 2000US-0249246P.
 PR 17-NOV-2000; 2000US-0249247P.
 PR 17-NOV-2000; 2000US-0249248P.
 PR 17-NOV-2000; 2000US-0249249P.
 PR 17-NOV-2000; 2000US-0249250P.
 PR 01-DEC-2000; 2000US-0250160P.
 PR 01-DEC-2000; 2000US-0250391P.
 PR 01-DEC-2000; 2000US-0251030P.
 PR 05-DEC-2000; 2000US-0251988P.
 PR 05-DEC-2000; 2000US-0256719P.
 PR 06-DEC-2000; 2000US-0251472P.
 PR 08-DEC-2000; 2000US-0251858P.
 PR 08-DEC-2000; 2000US-0251868P.
 PR 08-DEC-2000; 2000US-0251869P.
 PR 08-DEC-2000; 2000US-0251989P.
 PR 11-DEC-2000; 2000US-0251990P.
 PR 11-DEC-2000; 2000US-0254097P.
 PR 05-JAN-2001; 2001US-0259678P.
 PA (HUMA-) HUMAN GENOME SCI INC.
 PI Roen CA, Barash SC, Ruben SM;
 XX WPI; 2001-488783/53.
 DR P-PSDB; AAU16462.
 XX
 PT New nucleic acid molecules encoding 461 human secreted proteins for
 PT diagnosing, preventing, treating or ameliorating medical conditions and
 PT used as food additives or preservatives.
 XX
 PS Claim 1; SEQ ID NO 628; 980pp; English.
 XX
 CC The invention relates to isolated nucleic acid molecules and their
 CC encoded secreted proteins. The nucleic acids and proteins are used to
 CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,
 CC rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used
 CC in diagnosing a pathological condition or susceptibility to a
 CC pathological condition. Antibodies to the proteins can also be used in
 CC alleviating symptoms associated with the disorders and in diagnostic
 CC immunoassays e.g. radioimmunoassays or enzyme linked immunosorbent assays
 CC (ELISA). Disorders which are diagnosed or treated include autoimmune
 CC diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g.
 CC neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac
 CC arrest, cerebrovascular disorders e.g. cerebral ischaemia, angiogenesis,
 CC nervous system disorders e.g. Alzheimer's disease, infections caused by
 CC bacteria, viruses and fungi and ocular disorders e.g. corneal infection,
 CC and many other disorders listed in the specification. The polypeptides
 CC can also be used to aid wound healing and epithelial cell proliferation,
 CC to prevent skin aging due to sunburn, to maintain organs before
 CC transplantation, for supporting cell culture of primary tissues, to
 CC regenerate tissues and in chemotaxis. The polypeptides can also be used
 CC as a food additive or preservative to increase or decrease storage
 CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,
 CC minerals, cofactors and other nutritional components. The present
 CC sequence encodes a novel secreted protein of the invention. Note: The
 CC sequence data for this patent did not form part of the printed
 Query Match 60.0%; Score 773; DB 4; Length 2449;
 Best Local Similarity 100.0%; Pred. No. 6.7e-216;
 Matches 773; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 514 GTCTCCAGTGAAGTCTCAACTTTGGATTAAGAAAGCCGCGAGTCCCACTGGAATCCT 573
 DB 95 GTCTCCAGTGAAGTCTCAACTTTGGATTAAGAAAGCCGCGAGTCCCACTGGAATCCT 154
 QY 574 CAGCATCTTCTGATTTCTACAGCCAGTTTGGCTCGAGCTCTCCAGACATCAGAGGCG 633
 DB 155 CAGCATCTTCTGATTTCTACAGCCAGTTTGGCTCGAGCTCTCCAGACATCAGAGGCG 214
 QY 634 AATACCTGCTGCTGATCTGGGCCCAAGAGCGGTATGCAAAATCTTAAGTTTACGGGA 693
 XX |||||

DB 215 AATACCTGCTGCTGATCTGGGCCCAAGAGCGGTATGCAAAATCTCTAAGTTTACGGGA 274
 QY 694 CTCTCAATGACCACCTATCAGTCACTGGCTGGCCCAAGTCAAGTACCAGCTTAGGAAAAACG 753
 DB 275 CTCTCAATGACCACCTATCAGTCACTGGCTGGCCCAAGTCAAGTACCAGCTTAGGAAAAACG 334
 QY 754 GCGGGACAAAATTTCTGAAAAAATGAGCAAGCCCAACCCCATCTTTTATTGCAAGTAC 813
 DB 335 GCGGGACAAAATTTCTGAAAAAATGAGCAAGCCCAACCCCATCTTTTATTGCAAGTAC 394
 QY 814 TGTGCTCCCAAGTTTCAGAAACCCCTTCTACCTACATCAGTCACTTAGAATCTCACCTGGGT 873
 DB 395 TGTGCTCCCAAGTTTCAGAAACCCCTTCTACCTACATCAGTCACTTAGAATCTCACCTGGGT 454
 QY 874 TTCAAAATGAAGACATGACCCGCTTGTCACTGAGCAAGCAAGCAAGTGGAGCAAGAG 933
 DB 455 TTCAAAATGAAGACATGACCCGCTTGTCACTGAGCAAGCAAGCAAGTGGAGCAAGAG 514
 QY 934 ATCTCCCGGGTATCGTGGCTCAGAGGTCTCCAGAAACAAATAGCTGCCGAAAGAGACACA 993
 DB 515 ATCTCCCGGGTATCGTGGCTCAGAGGTCTCCAGAAACAAATAGCTGCCGAAAGAGACACA 574
 QY 994 GACTCTAAATTCAGTGTAAAGTTGTCTGCGACATTTGTGAGCAAAACATGCGGTAAAA 1053
 DB 575 GACTCTAAATTCAGTGTAAAGTTGTCTGCGACATTTGTGAGCAAAACATGCGGTAAAA 634
 QY 1054 CTCACCTTAAGCAAAACGACAGCAAGTCAACCCGCAACCACTTTCACAGTTTGTAAACAGAC 1113
 DB 635 CTCACCTTAAGCAAAACGACAGCAAGTCAACCCGCAACCACTTTCACAGTTTGTAAACAGAC 694
 QY 1114 GTGATGAAGAAATAGCTCTGAGGACGAATGCTTAGTTTCACCTTTCAGGCTGGATCC 1173
 DB 695 GTGATGAAGAAATAGCTCTGAGGACGAATGCTTAGTTTCACCTTTCAGGCTGGATCC 754
 QY 1174 CCTCAGCTGAACCTCTCTGTTGACCATCTCTGTTGACATTTGAACCTTGAACCT 1233
 DB 755 CCTCAGCTGAACCTCTCTGTTGACCATCTCTGTTGACATTTGAACCTTGAACCT 814
 QY 1234 CCTCCTGACACCTCGGCTCTGAGAGAGTGCACAAAAAATAAAAAAAAAAAT 1286
 DB 815 CCTCCTGACACCTCGGCTCTGAGAGAGTGCACAAAAAATAAAAAAAAAAAT 867
 RESULT 6
 ABX73790
 ID ABX73790 standard; DNA; 2449 BP.
 XX
 AC ABX73790;
 XX
 DT 18-MAR-2003 (first entry)
 XX
 DE Human novel polynucleotide #618.
 XX
 KW Human; gene; ds; neural disorder; immune system disorder; renal disorder;
 KW muscular disorder; respiratory disease; reproductive disorder;
 KW gastrointestinal disorder; pulmonary disorder; cardiovascular disorder;
 KW hyperproliferative disorder; inflammatory disease; allergic reaction;
 KW blood related disorder; cancer; immunosuppressive; antiinflammatory;
 KW cardiovascular; nephrotropic; cytostatic; antiallergic; thrombolytic;
 KW haemostatic; antiarteriosclerotic.
 XX
 OS Homo sapiens.
 XX
 PN US2002132753-A1.
 XX
 PD 19-SEP-2002.
 XX
 PF 17-JAN-2001; 2001US-00764864.
 XX
 PR 31-JAN-2000; 2000US-0179065P.
 PR 04-FEB-2000; 2000US-0180628P.
 PR 28-JUN-2000; 2000US-0214886P.
 PR 07-JUL-2000; 2000US-0216647P.

PR	07-JUL-2000;	2000US-0216880P.	CC	leukaemia), inflammatory diseases (e.g. septic shock, bursitis and
PR	11-JUL-2000;	2000US-0217487P.	CC	appendicitis), allergic reactions and conditions (e.g. asthma), blood
PR	11-JUL-2000;	2000US-0217496P.	CC	related disorders (e.g. thrombosis, atherosclerosis and myocardial
PR	14-JUL-2000;	2000US-0218290P.	CC	infarction) and cancerous diseases. Sequences ABX73173-ABX74167 represent
PR	26-JUL-2000;	2000US-0220963P.	CC	human novel polynucleotides of the invention
PR	26-JUL-2000;	2000US-0220964P.	XX	
PR	14-AUG-2000;	2000US-0224518P.	SQ	Sequence 2449 BP; 746 A; 526 C; 477 G; 687 T; 0 U; 13 Other;
PR	14-AUG-2000;	2000US-0224519P.		
PR	14-AUG-2000;	2000US-0225267P.		
PR	14-AUG-2000;	2000US-0225268P.		
PR	14-AUG-2000;	2000US-0225270P.		
PR	14-AUG-2000;	2000US-0225447P.		
PR	14-AUG-2000;	2000US-0225757P.		
PR	14-AUG-2000;	2000US-0225758P.		
PR	22-AUG-2000;	2000US-0226868P.		
PR	30-AUG-2000;	2000US-0228924P.		
PR	01-SEP-2000;	2000US-0229287P.		
PR	01-SEP-2000;	2000US-0229343P.		
PR	01-SEP-2000;	2000US-0229344P.		
PR	01-SEP-2000;	2000US-0229345P.		
PR	05-SEP-2000;	2000US-0229509P.		
PR	05-SEP-2000;	2000US-0229513P.		
PR	08-SEP-2000;	2000US-0231413P.		
PR	21-SEP-2000;	2000US-0234223P.		
PR	21-SEP-2000;	2000US-0234274P.		
PR	25-SEP-2000;	2000US-0234997P.		
PR	27-SEP-2000;	2000US-0235834P.		
PR	29-SEP-2000;	2000US-0236327P.		
PR	29-SEP-2000;	2000US-0236367P.		
PR	29-SEP-2000;	2000US-0236368P.		
PR	29-SEP-2000;	2000US-0236369P.		
PR	29-SEP-2000;	2000US-0236370P.		
PR	02-OCT-2000;	2000US-0236379P.		
PR	02-OCT-2000;	2000US-0236802P.		
PR	02-OCT-2000;	2000US-0237037P.		
PR	02-OCT-2000;	2000US-0237038P.		
PR	02-OCT-2000;	2000US-0237039P.		
PR	13-OCT-2000;	2000US-0237040P.		
PR	20-OCT-2000;	2000US-0240960P.		
PR	20-OCT-2000;	2000US-0241785P.		
PR	20-OCT-2000;	2000US-0241809P.		
PR	01-NOV-2000;	2000US-0244617P.		
PR	17-NOV-2000;	2000US-0249299P.		
PR	08-DEC-2000;	2000US-0251856P.		
PR	08-DEC-2000;	2000US-0251868P.		
PR	08-DEC-2000;	2000US-0251869P.		
XX				
PA	(ROSE/) ROSEN C A.			
PA	(RUBE/) RUBEN S M.			
PA	(BARA/) BARASH S C.			
PI	Rosen CA, Ruben SM, Barash SC;			
XX				
DR	WPI; 2003-147444/14.			
DR	P-PSDB; ABUS5530.			
XX				
PT	New polypeptides and nucleic acids, useful in gene therapy for treating,			
PT	inhibiting or preventing e.g. neural, immune system, muscular,			
PT	respiratory, reproductive, gastrointestinal, pulmonary, cardiovascular or			
PT	renal disorders.			
XX				
PS	Claim 1; SEQ ID NO 628; 402pp; English.			
XX				
CC	The invention relates to human novel polypeptides and their associated			
CC	polynucleotides. The polypeptides and polynucleotides are useful in gene			
CC	therapy for treating, inhibiting or preventing neural disorders, immune			
CC	system disorders (e.g. systemic lupus erythematosus, rheumatoid arthritis			
CC	and multiple sclerosis), muscular disorders, respiratory diseases (e.g.			
CC	nasal vestibulitis, nasal polyps and sinusitis), reproductive disorders,			
CC	gastrointestinal disorders, pulmonary disorders, cardiovascular disorders			
CC	(e.g. congenital heart defects, Ebstein's anomaly and hypoplastic left			
CC	heart syndrome), renal disorders (e.g. acute kidney failure and end-stage			
CC	renal disease), hyperproliferative disorders (e.g. Hodgkin's disease and			

CC	leukaemia), inflammatory diseases (e.g. septic shock, bursitis and
CC	appendicitis), allergic reactions and conditions (e.g. asthma), blood
CC	related disorders (e.g. thrombosis, atherosclerosis and myocardial
CC	infarction) and cancerous diseases. Sequences ABX73173-ABX74167 represent
CC	human novel polynucleotides of the invention
XX	
SQ	Sequence 2449 BP; 746 A; 526 C; 477 G; 687 T; 0 U; 13 Other;
	Query Match
	Best Local Similarity 60.0%; Score 773; DB 8; Length 2449;
	Matches 773; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	514 GTCTCAGTGAAGTCTCAACTTTTGATATAAAAGGCGGCGAGTCCAACTGGATCCT 573
DB	95 GTCTCAGTGAAGTCTCAACTTTTGATATAAAAGGCGGCGAGTCCAACTGGATCCT 154
QY	574 CAGCATCTTCTGATTTCTACAAGCCAGTTTGCTCGAGCTCTTCAGACATCAGAGGCG 633
DB	155 CAGCATCTTCTGATTTCTACAAGCCAGTTTGCTCGAGCTCTTCAGACATCAGAGGCG 214
QY	634 AATACCTGCTCTGATCTGAGCCCAAGAGCGGTATGCAAAATCTTAAGTTTACGGGA 693
DB	215 AATACCTGCTCTGATCTGAGCCCAAGAGCGGTATGCAAAATCTTAAGTTTACGGGA 274
QY	694 CTCTCAATGACCACATCAGTCACTGGCTGGCCAACTCAAGTACCAGCTTAGGAAAACG 753
DB	275 CTCTCAATGACCACATCAGTCACTGGCTGGCCAACTCAAGTACCAGCTTAGGAAAACG 334
QY	754 GCGGGACAAAATTTCTGAAAAACATGGAACAAGGCCACCCATCTTTTATTGCACTGAC 813
DB	335 GCGGGACAAAATTTCTGAAAAACATGGAACAAGGCCACCCATCTTTTATTGCACTGAC 394
QY	814 TGTGCTCCAGTTCAGAACCCCTTCTACCTACATCAGTCACTTAGAATCTCACCTGGGT 873
DB	395 TGTGCTCCAGTTCAGAACCCCTTCTACCTACATCAGTCACTTAGAATCTCACCTGGGT 454
QY	874 TTCCAAATGAAGACATGACCCGCTTGTCACTGGACCAAGCAAGAGTGGAGCAAGAG 933
DB	455 TTCCAAATGAAGACATGACCCGCTTGTCACTGGACCAAGCAAGAGTGGAGCAAGAG 514
QY	934 ATCTCCCGGGTATCGTGGCTCAGAGTCTCAGAAAACATAGTGTCCGAGAGGACACA 993
DB	515 ATCTCCCGGGTATCGTGGCTCAGAGTCTCAGAAAACATAGTGTCCGAGAGGACACA 574
QY	994 GACTCTAAATTCAGTGTAAAGTTGTGCTCGGACATTTGTGAGCAAAACATGCGGTAAA 1053
DB	575 GACTCTAAATTCAGTGTAAAGTTGTGCTCGGACATTTGTGAGCAAAACATGCGGTAAA 634
QY	1054 CTCCACCTAAGCAAAACGACAGCAAGTCAACCGAACACCATTTCAAGTTTGTAAACAGAC 1113
DB	635 CTCCACCTAAGCAAAACGACAGCAAGTCAACCGAACACCATTTCAAGTTTGTAAACAGAC 694
QY	1114 GTGGATGAAGAATAGCTCTGAGGACGAATGCTTAGTTTCCACTTTCCAGCTTGATCC 1173
DB	695 GTGGATGAAGAATAGCTCTGAGGACGAATGCTTAGTTTCCACTTTCCAGCTTGATCC 754
QY	1174 CTTCACTGTAACCTTCTTCTGTCAGCATCTCTGCTCTGACATTTGACATTTCACTTCACT 1233
DB	755 CTTCACTGTAACCTTCTTCTGTCAGCATCTCTGCTCTGACATTTGACATTTCACTTCACT 814
QY	1234 CTTCTGTGACACCTCTGCTCTGAGAGACCTGCCAAAAAATAAAAAAATAAAT 1286
DB	815 CTTCTGTGACACCTCTGCTCTGAGAGACCTGCCAAAAAATAAAAAAATAAAT 867
	RESULT 7
	AAS26014
ID	AAS26014 standard; cDNA; 799 BP.
XX	AAS26014;
AC	AAS26014;
XX	
DT	07-NOV-2001 (first entry)
XX	

DE Human cDNA encoding a novel secreted protein, Seq ID 193.
XX Human; immunosuppressive; antiarthritic; ss; antirheumatic; cytostatic;
KW cardiant; vasotropic; cerebroprotective; nootropic; neuroprotective;
KW antibacterial; virucide; fungicide; ophthalmological; vulnerary;
KW secreted protein; rheumatoid arthritis; hyperproliferative disorder;
KW cardiovascular disorder; cardiac arrest; cerebrovascular disorder;
KW cerebral ischaemia; angiogenesis; nervous system disorder;
KW Alzheimer's disease; infection; ocular disorder; corneal infection;
KW wound healing; epithelial cell proliferation; skin ageing; food additive;
KW preservative; antiproliferative.
XX Homo sapiens.
XX WO200153322-A2.
XX 02-AUG-2001.
XX 17-JAN-2001; 2001WO-US001341.
XX 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0218647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226686P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 06-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0233397P.
PR 14-SEP-2000; 2000US-0233398P.
PR 14-SEP-2000; 2000US-0233399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249246P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249267P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.

05-DEC-2000; 2000US-0251988P.
05-DEC-2000; 2000US-0256719P.
06-DEC-2000; 2000US-0251479P.
08-DEC-2000; 2000US-0251856P.
08-DEC-2000; 2000US-0251868P.
08-DEC-2000; 2000US-0251869P.
08-DEC-2000; 2000US-0251989P.
08-DEC-2000; 2000US-0251990P.
11-DEC-2000; 2000US-0254097P.
05-JAN-2001; 2001US-0259678P.
(HUMA-) HUMAN GENOME SCI INC.
Rosen CA, Barash SC, Ruben SM;
WPI; 2001-488783/53.
P-PSDB; AAU16027.
New nucleic acid molecules encoding 461 human secreted proteins for
diagnosing, preventing, treating or ameliorating medical conditions and
used as food additives or preservatives.
Claim 1; SEQ ID NO 193; 980pp; English.
The invention relates to isolated nucleic acid molecules and their
encoded secreted proteins. The nucleic acids and proteins are used to
prevent, treat or ameliorate a medical condition in e.g. humans, mice,
rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used
in diagnosing a pathological condition or susceptibility to a
pathological condition. Antibodies to the proteins can also be used in
alleviating symptoms associated with the disorders and in diagnostic
immunoassays e.g. radioimmunoassays or enzyme linked immunosorbent assays
(ELISA). Disorders which are diagnosed or treated include autoimmune
diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g.
neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac
arrest, cerebrovascular disorders e.g. cerebral ischaemia, angiogenesis,
nervous system disorders e.g. Alzheimer's disease, infections caused by
bacteria, viruses and fungi and ocular disorders e.g. corneal infection,
and many other disorders listed in the specification. The polypeptides
can also be used to aid wound healing and epithelial cell proliferation,
to prevent skin aging due to sunburn, to maintain organs before
transplantation, for supporting cell culture of primary tissues, to
regenerate tissues and in chemotaxis. The polypeptides can also be used
as a food additive or preservative to increase or decrease storage
capabilities, fat content, lipid, protein, carbohydrate, vitamins,
minerals, cofactors and other nutritional components. The present
sequence encodes a novel secreted protein of the invention. Note: The
sequence data for this patent did not form part of the printed
Query Match 59.8%; Score 770.4; DB 4; Length 799;
Best Local Similarity 99.9%; Pred. No. 2.2e-215;
Matches 771; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
514 GTCTCCAGTGAAGTCTCAACTTTGTCATAAAAGAAAAGCGCGGAGTCCAACTGGAATCCT 573
12 GTCTCCAGTGAAGTCTCAACTTTGTCATAAAAGAAAAGCGCGGAGTCCAACTGGAATCCT 71
574 CAGCATCTCTGATTTACAGCCCGAGTTTGCTCGAGCCTCTTCAGACATCAGAGGCG 633
72 CAGCATCTCTGATTTACAGCCCGAGTTTGCTCGAGCCTCTTCAGACATCAGAGGCG 131
634 AATACTCTGCTGCTGATCTGGCCGACAGAGCGGTATGCAATCTCTAAGTTTACGGGA 693
132 AATACTCTGCTGCTGATCTGGCCGACAGAGCGGTATGCAATCTCTAAGTTTACGGGA 191
694 CTCTCAATGACCACTATCAGTCACTGGCTGGCCCAAGCTCAAGTACCAGCTTTAGGAAAACG 753
192 CTCTCAATGACCACTATCAGTCACTGGCTGGCCCAAGCTCAAGTACCAGCTTTAGGAAAACG 251
754 GCGGGGACAAAATTTCTGAAAAAATGAGCAAAAGGCGACCCCATCTTTATTCAGTGAC 813
252 GCGGGGACAAAATTTCTGAAAAAATGAGCAAAAGGCGACCCCATCTTTATTCAGTGAC 911

QY 814 TGTGCTCCAGTTCAGAACCCCTTCTACATCATCAGTCACTTAAATCTCACCTGGGT 873
DB 312 TGTGCTCCAGTTCAGAACCCCTTCTACATCATCAGTCACTTAAATCTCACCTGGGT 371
QY 874 TTCCAATGAAGGACATGACCCGCTTGTCACTGGACCGAAGCAAGGTGAGCAAGAG 933
DB 372 TTCCAATGAAGGACATGACCCGCTTGTCACTGGACCGAAGCAAGGTGAGCAAGAG 431
QY 934 ATCTCCCGGGTATCGTCCGCTCAGAGGTCTCCAGAAAACAATAGCTGCCGAGGACACA 993
DB 432 ATCTCCCGGGTATCGTCCGCTCAGAGGTCTCCAGAAAACAATAGCTGCCGAGGACACA 491
QY 994 GACTCTAAATTCAGTGTAAAGTTGTCTGCGACATTTGTGAGCAAAACATCGGTAAAA 1053
DB 492 GACTCTAAATTCAGTGTAAAGTTGTCTGCGACATTTGTGAGCAAAACATCGGTAAAA 551
QY 1054 CTCCACCTTAAGCAAAACGACAGCAAGTCAACCGGACACCATTCACAGTTTGTAAACAG 1113
DB 552 CTCCACCTTAAGCAAAACGACAGCAAGTCAACCGGACACCATTCACAGTTTGTAAACAG 611
QY 1114 GTGGATGAAGATAGCTCTGCGAGGACGAATGCTTAGTTTCCACTTTCAGGCTGGATCC 1173
DB 612 GTGGATGAAGATAGCTCTGCGAGGACGAATGCTTAGTTTCCACTTTCAGGCTGGATCC 671
QY 1174 CTTCACTGAACCCCTTCTTCTGCGACCATCTCTGACATTTGAATCACTTGAAT 1233
DB 672 CTTCACTGAACCCCTTCTTCTGCGACCATCTCTGCGACATTTGAATCACTTGAAT 731
QY 1234 CTTCTGACACCTGGCTCTGAGAGACTGCGCAAAAAAAGGCAAAAAA 1285
DB 732 CTTCTGACACCTGGCTCTGAGAGACTGCGCAAAAAAAGGCAAAAAA 783
RESULT 8
ABX73355
ID ABX73355 standard; DNA; 799 BP.
AC ABX73355;
XX
DT 18-MAR-2003 (first entry)
XX Human novel polynucleotide #183.
DE
XX Human; gene; ds; neural disorder; immune system disorder; renal disorder;
muscular disorder; respiratory disease; reproductive disorder;
gastrointestinal disorder; pulmonary disorder; cardiovascular disorder;
hyperproliferative disorder; inflammatory disease; allergic reaction;
blood related disorder; cancer; immunosuppressive; antinflammatory;
cardiovascular; nephrotropic; cytostatic; antiallergic; thrombolytic;
haemostatic; antiarteriosclerotic.
XX
XX Homo sapiens.
OS
XX US2002132753-A1.
PN
XX
PD 19-SEP-2002.
XX
PF 17-JAN-2001; 2001US-00764864.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 28-JUN-2000; 2000US-0214886P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.

PR 14-AUG-2000; 2000US-0225270P.
 PR 14-AUG-2000; 2000US-0225447P.
 PR 14-AUG-2000; 2000US-0225757P.
 PR 14-AUG-2000; 2000US-0225758P.
 PR 22-AUG-2000; 2000US-0226868P.
 PR 30-AUG-2000; 2000US-0228924P.
 PR 01-SEP-2000; 2000US-0229287P.
 PR 01-SEP-2000; 2000US-0229343P.
 PR 01-SEP-2000; 2000US-0229344P.
 PR 01-SEP-2000; 2000US-0229345P.
 PR 05-SEP-2000; 2000US-0229509P.
 PR 05-SEP-2000; 2000US-0229513P.
 PR 08-SEP-2000; 2000US-0231413P.
 PR 21-SEP-2000; 2000US-0234223P.
 PR 21-SEP-2000; 2000US-0234274P.
 PR 25-SEP-2000; 2000US-0234997P.
 PR 27-SEP-2000; 2000US-0235834P.
 PR 29-SEP-2000; 2000US-0236327P.
 PR 29-SEP-2000; 2000US-0236367P.
 PR 29-SEP-2000; 2000US-0236368P.
 PR 29-SEP-2000; 2000US-0236369P.
 PR 29-SEP-2000; 2000US-0236370P.
 PR 02-OCT-2000; 2000US-0236802P.
 PR 02-OCT-2000; 2000US-0237037P.
 PR 02-OCT-2000; 2000US-0237038P.
 PR 02-OCT-2000; 2000US-0237039P.
 PR 02-OCT-2000; 2000US-0237040P.
 PR 13-OCT-2000; 2000US-0239935P.
 PR 20-OCT-2000; 2000US-0240960P.
 PR 20-OCT-2000; 2000US-0241785P.
 PR 20-OCT-2000; 2000US-0241809P.
 PR 01-NOV-2000; 2000US-0244617P.
 PR 17-NOV-2000; 2000US-0249299P.
 PR 08-DEC-2000; 2000US-0251856P.
 PR 08-DEC-2000; 2000US-0251869P.
 PR 08-DEC-2000; 2000US-0251869P.
 XX
 PA (ROSE/) ROSEN C A.
 PA (RUBE/) RUBEN S M.
 PA (BARA/) BARASH S C.
 PI Rosen CA, Ruben SM, Barash SC;
 XX
 DR WPI; 2003-147444/14.
 DR P-PSDB; ABU55095.
 XX
 PT New polypeptides and nucleic acids, useful in gene therapy for treating,
 PT inhibiting or preventing e.g. neural, immune system, muscular,
 PT respiratory, reproductive, gastrointestinal, pulmonary, cardiovascular or
 PT renal disorders.
 XX
 PS Claim 1; SEQ ID NO 193; 402pp; English.
 PS
 XX The invention relates to human novel polypeptides and their associated
 CC polynucleotides. The polypeptides and polynucleotides are useful in gene
 CC therapy for treating, inhibiting or preventing neural disorders, immune
 CC system disorders (e.g. systemic lupus erythematosus, rheumatoid arthritis
 CC and multiple sclerosis), muscular disorders, respiratory diseases (e.g.
 CC nasal vestibulitis, nasal polyps and sinusitis), reproductive disorders,
 CC gastrointestinal disorders, pulmonary disorders, cardiovascular disorders
 CC (e.g. congenital heart defects, Epstein's anomaly and hypoplastic left
 CC heart syndrome), renal disorders (e.g. acute kidney failure and end-stage
 CC renal disease), hyperproliferative disorders (e.g. Hodgkin's disease and
 CC leukaemia), inflammatory diseases (e.g. septic shock, bursitis and
 CC appendicitis), allergic reactions and conditions (e.g. asthma), blood
 CC related disorders (e.g. thrombosis, atherosclerosis and myocardial
 CC infarction) and cancerous diseases. Sequences ABX7131-ABX74167 represent
 CC human novel polynucleotides of the invention
 XX
 SQ Sequence 799 BP; 247 A; 218 C; 162 G; 172 T; 0 U; 0 Other;
 Query Match 59.8%; Score 770.4; DB 8; Length 799;
 Best Local Similarity 99.9%; Pred. NO. 2.2e-215;

Matches 771; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 514 GTCTCCAGTGAAGTCTCAACTTTTCATATAAAGAAAGCCGGCAGTCCAACTGGGAATCCT 573
 |||||
 DB 12 GTCTCCAGTGAAGTCTCAACTTTTCATATAAAGAAAGCCGGCAGTCCAACTGGGAATCCT 71
 |||||
 QY 574 CAGCATCTTCTGATTTCTACAAGCCAGTTTGCCTCGAGCCTCTTCCAGACATCAGAGGC 633
 |||||
 DB 72 CAGCATCTTCTGATTTCTACAAGCCAGTTTGCCTCGAGCCTCTTCCAGACATCAGAGGC 131
 |||||
 QY 634 AATACCTGCTGTCTGATCTGGGCCCAACAAGAGGTATGCAAAATCTCTAAGTTTACGGGA 693
 |||||
 DB 132 AATACCTGCTGTCTGATCTGGGCCCAACAAGAGGTATGCAAAATCTCTAAGTTTACGGGA 191
 |||||
 QY 694 CTCTCAATGACCACTATACAGTCACTGGCTGGCCAAAGTCAAGTACCAGCTTAGGAAAACG 753
 |||||
 DB 192 CTCTCAATGACCACTATACAGTCACTGGCTGGCCAAAGTCAAGTACCAGCTTAGGAAAACG 251
 |||||
 QY 754 GCGGGACAAAATTTCTGAAAACATGACAAAGGCCACCCCATCTTTTATTGCAAGTAC 813
 |||||
 DB 252 GCGGGACAAAATTTCTGAAAACATGACAAAGGCCACCCCATCTTTTATTGCAAGTAC 311
 |||||
 QY 814 TGTGCTCTCCAGTTTCAAGAACCCCTTCTACCTACATCAGTCACTTAGAATCTCACCTGGGT 873
 |||||
 DB 312 TGTGCTCTCCAGTTTCAAGAACCCCTTCTACCTACATCAGTCACTTAGAATCTCACCTGGGT 371
 |||||
 QY 874 TTCAAATGAAGAGCATGACCCGCTTGTCACTGGACCAAGCAAGAGTGGAGCAAGAG 933
 |||||
 DB 372 TTCAAATGAAGAGCATGACCCGCTTGTCACTGGACCAAGCAAGAGTGGAGCAAGAG 431
 |||||
 QY 934 ATCTCCCGGGTATCGTGGCTCAGAGTCTCCAGAAACAATAGTCTGCCGAGAGAGACACA 993
 |||||
 DB 432 ATCTCCCGGGTATCGTGGCTCAGAGTCTCCAGAAACAATAGTCTGCCGAGAGAGACACA 491
 |||||
 QY 994 GACTCTAAATTTCAAGTGTAAAGTGTGCTGCGGACATTTGTGAGCAACATCGCGTAAAA 1053
 |||||
 DB 492 GACTCTAAATTTCAAGTGTAAAGTGTGCTGCGGACATTTGTGAGCAACATCGCGTAAAA 551
 |||||
 QY 1054 CTCACCTTAAGCAAAACGACAGCAAGTCAACCGAACAACATTACAGTTTGTAAACAGAC 1113
 |||||
 DB 552 CTCACCTTAAGCAAAACGACAGCAAGTCAACCGAACAACATTACAGTTTGTAAACAGAC 611
 |||||
 QY 1114 GTGATGAAGAAATAGTCTGAGGACGAATGCTTTCACCTTTCCAGCTGGATCC 1173
 |||||
 DB 612 GTGATGAAGAAATAGTCTGAGGACGAATGCTTTCACCTTTCCAGCTGGATCC 671
 |||||
 QY 1174 CCTCAGTGAACCTCTTCTCGTTGGCACCATCTGCTTCTGACATTGAACCTATTGAACCT 1233
 |||||
 DB 672 CCTCAGTGAACCTCTTCTCGTTGGCACCATCTGCTTCTGACATTGAACCTATTGAACCT 731
 |||||
 QY 1234 CCTCCTGACACCTCTGAGTCTGAGAGAGTGCACAAAAAAGAAAAAAGAAAAA 1285
 |||||
 DB 732 CCTCCTGACACCTCTGAGTCTGAGAGAGTGCACAAAAAAGAAAAAAGAAAAA 783
 |||||
 RESULT 9
 AAI24344
 ID AAI24344 standard; DNA; 2012 Bp.
 AC AAI24344;
 XX
 XX 12-OCT-2001 (first entry)
 DT
 XX
 XX Probe #14277 for gene expression analysis in human cervical cell sample.
 DB
 XX Probe; human; microarray; gene expression; cervical epithelial cell;
 KW cervical cancer; ss.
 KW
 XX Homo sapiens.
 OS
 XX WO200157278-A2.
 EN
 XX PD 09-AUG-2001.


```
Db 1504 TCGGCGCTCGCAACACGCCCCCGGCGCTGCATGATCAACCCACTAGCGCCCTGCGAG 1563
Qy 121 TCCGTCCTGAAACAATCACTTGGGCAAGCCACGAGCCCTTGGCTCACCTTCTGCTCC 180
Db 1564 TCCGTCCTGAAACAATCACTTGGGCAAGCCACGAGCCCTTGGCTCACCTTCTGCTCC 1623
Qy 181 AGCCCAAGTTCAAGCACAAATTTCCATGTTCCACAAGTCCGAATCTCAATGTCATGACAAG 240
Db 1624 AGCCCAAGTTCAAGCACAAATTTCCATGTTCCACAAGTCCGAATCTCAATGTCATGACAAG 1683
Qy 241 CCGGTCCTGAGTCCTGCTCCACAAAGGTCAGCCAGCGTGTCCAGCGCTTACCTGTTTCAG 300
Db 1684 CCGGTCCTGAGTCCTGCTCCACAAGGTCAGCCAGCGTGTCCAGCGCTTACCTGTTTCAG 1743
Qy 301 AACACGATCAGCCCAATGACCTGACCAAGTCCAAAAGCAAGAACCGGAGTCTCTCGCAA 360
Db 1744 AACACGATCAGCCCAATGACCTGACCAAGTCCAAAAGCAAGAACCGGAGTCTCTCGCAA 1803
Qy 361 GCACAATCTTGTATGTCCTCCACCTCAGAAAGCAAGCGTCTGTGTGACATCGCGACATGGTC 420
Db 1804 GCACAATCTTGTATGTCCTCCACCTCAGAAAGCAAGCGTCTGTGTGACATCGCGACATGGTC 1863
Qy 421 AAAGTCCTCCCAAGCCACACCCCAAGCCAGCGTCTCTCTCCAGGTCCTCCCGCCCATG 480
Db 1864 AAAGTCCTCCCAAGCCACACCCCAAGCCAGCGTCTCTCTCCAGGTCCTCCCGCCCATG 1923
Qy 481 AAGCTGGAATGGATGTGAGGCGCTTTGAGGATGTCTCCAG 521
Db 1924 AAGCTGGAATGGATGTGAGGCGCTTTGAGGATGTCTCCAG 1964

RESULT 11
AAI62638
ID AAI62638 standard; DNA; 3604 BP.
XX
AC AAI62638;
XX
DT 19-OCT-2001 (first entry)
DE Human breast or ovarian antigen genomic DNA SEQ ID NO: 288.
KW Human; breast antigen; ovarian antigen; cancer; metastasis; gene therapy;
KW ds.
XX
XX Homo sapiens.
XX
XX WO200155324-A2.
XX
PD 02-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US001344.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
```

08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249246P.
PR 17-NOV-2000; 2000US-0249247P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
(HUMA-) HUMAN GENOME SCI INC.
Rosen CA, Barash SC, Ruben SM;
WPI; 2001-488785/53.
New isolated nucleic acids and polypeptides, useful for diagnosing,
treating and/or preventing human diseases and disorders.
Disclosure; SEQ ID NO 288; 520pp + Sequence Listing; English.
The present invention provides the protein and coding sequences of a
number of ovarian and breast antigens. These are shown in AAI62467-
AAI62572 and AAM42240-AA42345. The sequences can be used in the
diagnosis, prevention and treatment of breast and ovarian cancers, and
their metastases. The present sequence is a genomic sequence of the
invention. Note: The sequence data for this patent did not form part
of the printed specification, but was obtained in electronic format directly
from WIPO at ftp.wipo.int/pub/published_pct_sequences
Sequence 3604 BP; 1002 A; 955 C; 853 G; 793 T; 0 U; 1 Other;
Query Match 30.3%; Score 390.4; DB 4; Length 3604;
Best Local Similarity 62.4%; Pred. No. 2.1e-103;
Matches 673; Conservative 0; Mismatches 381; Indels 24; Gaps 3;
52 AGCAATGGGTGGCCCTCGCCCAACACGCGCCCGCCCTGCGCATGCAATCAACCCACTCAGC 111
1243 AACAACTGGGGATCATCATGGACCACTACCGGAGCCTTCTTTCATCAACCCGCTGAGC 1302

QY 112 GCGCTGAGTCCGTCCTGAAACAATCACTTGGGCAAGCCACGAGACCTTGGCTCACCT 171
DB 1303 GCTTTGAGTCCATCATGAACACACCCAGCTGGGCAAGGTGTCCAAGCCCGTG----- 1353
QY 172 TCCTGCTCAGCCCAAGTTCAAGCAAAATTTCCATGTTTCCACAAGTCGATCTCAATGTC 231
DB 1354 -----AGTCCCTCGCTGGACCCGCTGGCGATGCTGTACAAAGATCAGCAACACGATG 1404
QY 232 ATGGACAAGCCGGTCTTGAGTCTGCTCCCAAGAGTCCAGCCAGCGGTGTCCAGGCGCTAC 291
DB 1405 CTGGACAAGCCGGTGTATCCCGCCACCCCTGTGAAGCAGGCGCGATGCCATCGACCGCTAC 1464
QY 292 CTGTTTGAGAACAGCGATCAGCCCATTCGACCAAGTCCAAAGTCCAAAGCAAGAAAGCCGAG 351
DB 1465 TATTATGAAACAGCGACCGCCCATTGACTTAACCAAGTCCAAAGCAAGAAAGCCGCTGGTG 1524
QY 352 TCCTGCAAGCAAAATCTTTGATGTCCCACTCAGAGCAGCTCTGTCTGACATCGCC 411
DB 1525 TCCAGCGTGGTGTATTCGGTGGCATCAGCTCTGCGGGAGAGCGCACTCATGACATCTCC 1584
QY 412 GACATGTTCAAGTCTCTCCCAAGCCACCCCAAGCCAGCCTCTCTCCAGGGTGC 471
DB 1585 GACATGGTGAACAACTCAGGCGCGCTGACGCCCAAGTCTCCAGCCCTCCACAGTT 1644
QY 472 CCCCCCATGAAGCTGGAATGGATGTGAGGCGCTTTGAGGATGTCTCCAGTGAAGTCTCA 531
DB 1645 TCAGA--GAAGTCCGATGCTGATGCGAGCAGCTTTGAGGAGCGTGGAGAGTGTCA 1701
QY 532 ACTTTGCAATAAAGAAAGCCGCGAGTCCAACTGGAATCTCTCAGCATCTTCTGATTCTA 591
DB 1702 CCGGTCCCAAGAGGAAGGCGCGCAGTCCAACTGGAACCCCGCAGCACCTTCTCATCTCG 1761
QY 592 CAAAGCCAGTTTGGCTCGAGCCTCTTCCAGACATCAGAGGGCAAAATACCTGTCTGAT 651
DB 1762 CAGGCCAGTTCGCTCGAGCTTGGGGAGACCAAGAGGGCAAGTACATCATGTGCGAC 1821
QY 652 CTGGGCCCCACAGAGCGTATGCAAAATCTCTAAGTTTACGGGACTCTCAATGACCACTATC 711
DB 1822 TTGGGCCCCAGAGAGGGTGCACATCTCGAAGTTTACTGGGCTCTTCCATGACCACTC 1881
QY 712 AGTCACTGCTGGCCAAAGTCAAGTACCAAGTTAGAAAACGGGGGGGAGCAAAATTTCTG 771
DB 1882 AGCCACTGGCTGGCCAAATGTGAAGTACCAGTTGAGAGGACAGGGGGGAAACGAAATTCCTA 1941
QY 772 AAAAAATGGCAAAAGGCCACCCCATCTTTTATTCAGTGAAGTGTGCTCCCTCCAGTTCAGA 831
DB 1942 AAGAACCTGGACACAGGGGATCTCTGTTTCTTTTGAACGATGTGCTCTCAGTTTCA 2001
QY 832 ACCCTTCTTACCTACATCAGTCACTTAGAATCTCACCTGGGTTTCCAAATGAAGGACATG 891
DB 2002 ACTGCTTCTACATACATAAGTCAATTTGGAGACACACTTGGGCTTACGCTGAAGGATCTC 2061
QY 892 ACCGCTTCTAGTGGACCAAGCAAGGTGGAGCAAGATCTCCCGGGTATCGTCG 951
DB 2062 TCCAAAGTCCCACT---CAATCAGATTCAAGAACACAGCAGAAATGTTTCGAAAGCTCCACC 2118
QY 952 GCTCAGAGTCTCCAGAAACAATAGCTGCCGAGAGGACACAGACTCTATAATTCAGATGT 1011
DB 2119 AACAAACTCTCTGGGCCCATCTGGGGGCCACCGAGGAAGACTTGGGCTTCCATTTCCAATGT 2178
QY 1012 AAGTTGTGCTGTGGGACATTTGTGAGCAAAACATGCGGTAAAACTCCACCTTAAGCAAAACG 1071
DB 2179 AAGCTCTCAACCGGACTTTTTCGAGCAAGCAGCAGTCAAACTGACCTTAGTAAGACC 2238
QY 1072 CACAGCAAGTACCCGAAACACCATTCACAGTTTGTAAACAGAGCTGATGAAGATAGC 1129
DB 2239 CACGGCAAGTCTCCCGAGGACCACTGATCTATGTGACTGAGTTGAGAAACAGTAGC 2296
RESULT 12
AAL06732
ID AAL06732 standard; DNA; 3604 BP.

XX AAL06732;
AC 14-SEP-2000; 2000US-0232400P.
XX 14-SEP-2000; 2000US-0232401P.
DT 14-SEP-2000; 2000US-0233063P.
XX 14-SEP-2000; 2000US-0233064P.
XX 14-SEP-2000; 2000US-0233065P.
DE 21-SEP-2000; 2000US-0234223P.
XX 21-SEP-2000; 2000US-0234274P.
XX 25-SEP-2000; 2000US-0234977P.
KW 25-SEP-2000; 2000US-0234998P.
XX 26-SEP-2000; 2000US-0235484P.
OS 27-SEP-2000; 2000US-0235834P.
XX 27-SEP-2000; 2000US-0235836P.
PN 29-SEP-2000; 2000US-0236327P.
XX 29-SEP-2000; 2000US-0236367P.
PD 29-SEP-2000; 2000US-0236368P.
XX 29-SEP-2000; 2000US-0236369P.
XX 29-SEP-2000; 2000US-0236370P.
XX 02-OCT-2000; 2000US-0236802P.
XX 02-OCT-2000; 2000US-0237037P.
XX 02-OCT-2000; 2000US-0237038P.
XX 02-OCT-2000; 2000US-0237039P.
XX 13-OCT-2000; 2000US-0237040P.
XX 13-OCT-2000; 2000US-0239935P.
XX 13-OCT-2000; 2000US-0239937P.
XX 20-OCT-2000; 2000US-0240960P.
XX 20-OCT-2000; 2000US-0241221P.
XX 20-OCT-2000; 2000US-0241785P.
XX 20-OCT-2000; 2000US-0241786P.
XX 20-OCT-2000; 2000US-0241787P.
XX 20-OCT-2000; 2000US-0241808P.
XX 20-OCT-2000; 2000US-0241809P.
XX 20-OCT-2000; 2000US-0241826P.
XX 01-NOV-2000; 2000US-0244617P.
XX 08-NOV-2000; 2000US-0246474P.
XX 08-NOV-2000; 2000US-0246476P.
XX 08-NOV-2000; 2000US-0246477P.
XX 08-NOV-2000; 2000US-0246478P.
XX 08-NOV-2000; 2000US-0246523P.
XX 08-NOV-2000; 2000US-0246524P.
XX 08-NOV-2000; 2000US-0246525P.
XX 08-NOV-2000; 2000US-0246526P.
XX 08-NOV-2000; 2000US-0246527P.
XX 08-NOV-2000; 2000US-0246528P.
XX 08-NOV-2000; 2000US-0246532P.
XX 08-NOV-2000; 2000US-0246609P.
XX 08-NOV-2000; 2000US-0246610P.
XX 08-NOV-2000; 2000US-0246611P.
XX 08-NOV-2000; 2000US-0246613P.
XX 17-NOV-2000; 2000US-0249207P.
XX 17-NOV-2000; 2000US-0249208P.
XX 17-NOV-2000; 2000US-0249209P.
XX 17-NOV-2000; 2000US-0249210P.
XX 17-NOV-2000; 2000US-0249211P.
XX 17-NOV-2000; 2000US-0249212P.
XX 17-NOV-2000; 2000US-0249213P.
XX 17-NOV-2000; 2000US-0249214P.
XX 17-NOV-2000; 2000US-0249215P.
XX 17-NOV-2000; 2000US-0249216P.
XX 17-NOV-2000; 2000US-0249217P.
XX 17-NOV-2000; 2000US-0249218P.
XX 17-NOV-2000; 2000US-0249244P.
XX 17-NOV-2000; 2000US-0249245P.
XX 17-NOV-2000; 2000US-0249264P.
XX 17-NOV-2000; 2000US-0249265P.
XX 17-NOV-2000; 2000US-0249297P.
XX 17-NOV-2000; 2000US-0249299P.
XX 17-NOV-2000; 2000US-0249300P.
XX 01-DEC-2000; 2000US-0250160P.
XX 01-DEC-2000; 2000US-0250391P.
XX 05-DEC-2000; 2000US-0251030P.
XX 05-DEC-2000; 2000US-0251988P.
XX 05-DEC-2000; 2000US-0256719P.

XX 21-NOV-2001 (first entry)
DE Human reproductive system related antigen DNA SEQ ID NO: 9420.
KW Human; reproductive system related antigen; reproductive system disorder;
XX cancer; gene therapy; ds.
OS Homo sapiens.
PN WO200153320-A2.
XX 02-AUG-2001.
XX 17-JAN-2001; 2001WO-US001339.
XX 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225216P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227003P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229503P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.

PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-465570/50.
XX
XX Isolated nucleic acid molecule encoding a reproductive system antigen is
XX used in preventing, treating or ameliorating a medical condition.
XX
XX Disclosure; SEQ ID NO 9420; 1297pp + Sequence Listing; English.
XX
XX The present invention provides the protein and coding sequences of a
XX number of human reproductive system related antigens. These can be used
XX in the prevention and treatment of reproductive system disorders,
XX including cancer. The present sequence is a genomic sequence encoding a
XX protein of the invention
XX
XX Sequence 3604 BP; 1002 A; 955 C; 853 G; 793 T; 0 U; 1 Other;
XX
Query Match 30.3%; Score 390.4; DB 4; Length 3604;
Best Local Similarity 62.4%; Pred. No. 2.1e-103;
Matches 673; Conservative 0; Mismatches 381; Indels 24; Gaps 3;
QY 52 AGCAATGGTGGCGCTCGCCACACACCGCCCGCCCTGCCATGATCAACCCACTCAGC 111
DB 1243 AACAACCTGGGATCATCATGACCACTCACCAGAGCCTTCTTCAATCAACCGCTGAC 1302
QY 112 GGCCTGCAGTCGTCTCGAACATCACTTTGGGCAAGCCACGAGCCCTTGGCGTCACT 171
DB 1303 GCTTTGCACTCCATCATGAACACCCACCTGGGCAAGGTGTCCAAGCCCGTG----- 1353
QY 172 TCGTGTCCAGCCCAAGTTCAAGCACAATTTTCATGTTCCACAAAGTCGAATCTCAATGTC 231
DB 1354 -----AGTCCCTCGCTGGACCCGCTGGCGATGCTGTACAAGATCAGCAACAGCATG 1404
QY 232 ATGGACAAGCGGTCTTGTAGTCTGCTCCACAAGTTCAGCCAGCGTGTCCAGGCGCTAC 291
DB 1405 CTGGACAGCCGGTGACCCGCCACCTGTGTGAACAGCGCGATGCCATCAACCGCTAC 1464
QY 292 CTGTTTGAACAAGCGATCAGCCCATTTGACCTGACCAAGTTCCAAAAGCAAGAAAGCCGAG 351
DB 1465 TATTATGAACAACAGCGACAGCCCATTTGACTTAACCAAGTTCCAAGAAACGCGCTGGTG 1524
QY 352 TCGTCCGAAGCACAATCTTGTATGTATGTCCTCCACCTCAGAAGCAGCTCTGTGTACATCGGC 411
DB 1525 TCCAGCGTGGCTGATTCGGTGGCATCACCTCTCGGGGAGAGCGCACTCATGACATCTCC 1584
QY 412 GACATGTCAAAGTCTCTCCCAAGCAACACCCCAAGCCAGCTCTCTCCAGGTC 471
DB 1585 GACATGTGAACAACTCTCAGCGCGCTGACGCCCAAGTCTCCAGCGCTCCACAGTT 1644
QY 472 CCCCCATGAAGCTGGAATGGAATGT CAGCGCTTTTGAGGATGTCTCCAGTGAAGTCTCA 531
DB 1645 TCAGA---GAAGTCCGATGCTGATGGCAGCAGCTTTTGAGGAGCGTTGGACGAGTGCA 1701
QY 532 ACTTTCATAAAGAAAGCCGGAGTCCAACTGGAATCTCAGCATCTTCTGTATCTTA 591
DB 1702 CCGGTCCACAAGAGGAGGCGCGCAGTCCAACTGGAACCCGCGAGCACCTTCTCATCTG 1761
QY 592 CAAAGCCAGTTTCCCTCGAGCCTTCCAGACATCAGAGGCAAAATACCTGCTGTGAT 651
DB 1762 CAGGCCAGTTTCGCTCGAGCTTTCGGGGAGACCAACAGAGGGCAAGTACATCATGTCGAC 1821
QY 652 CTGGGCCCCACAGAGCGTATGCAAAATCTCTAAGTTTACGGGACTCTCAATGACCCTATC 711

DB 1822 TTGGCCCGCAGGAGAGGGTGCACATCTCGAAGTTTACTGGGCTCTCCATGACCACCATC 1881
QY 712 AGTCACTGGCTGGCCAAAGTCAAGTACCAGCTTAGGAAAAACGGCGGGAACAAAATTTCTG 771
DB 1882 AGCCACTGGCTGGCCAAATGTGAAGTACCAGTTGAGAGGACAGGGGGAAACGAATTCCTA 1941
QY 772 AAAAATCATGGACAAAGGCCACCCCATCTTTTATTGCAGTACTGTGCCTCCAGTTTCA 831
DB 1942 AAGAACCTGGACACAGGGCATCTCTGTTTCTTTTCAACGATTGTGCTCTCAGTTTCA 2001
QY 832 ACCCTTCTACCTACATCAGTCACTTAGAATCTCAGCTGGGTTTCAATGAAGACATG 891
DB 2002 ACTGCTTCTACATACATAAGTCAATTTGGAGACACACTTGGGCTTCAAGGATCTC 2061
QY 892 ACCCGCTTGTAGTGGACAGCAAGCAAGGTGGAGAGAGAGATCTCCCGGTATCGTCG 951
DB 2062 TCAAAGCTGCCACT---CAATCAGATTCAAGAACAGCAGAAATGTTTCGAAAGTCTCCACC 2118
QY 952 GCTCAGAGTCTCCAGAAACAAATAGTCTGCCAAGAGGACACAGACTCTAAATTCAGTGT 1011
DB 2119 AACAAAATCTCGGGCCCACTGGGGCCACCGAGGAAGACTTGGGCTCCACATTTCCAATGT 2178
QY 1012 AAGTTGTGCTGCGACATTTGTGAGCAAAACATGCGGTAAAACTCCACCTAAGCAAAAGC 1071
DB 2179 AAGCTCTGCACCGGACTTTTTCGAGCAAGCAGCAGTCAAACTGCACCTTAGTAAGACC 2238
QY 1072 CACAGCAAGTCAACCCGAAACCAATTCACAGTTTGTAAACAGACGTGATCAAGAAATAGC 1129
DB 2239 CACGCAAGTCTCCCGAGGACCACCTGATCTATGTGACTGAGTTGGAGAAACAGTAGC 2296
RESULT 13
ABK34640
ID ABK34640 standard; cDNA; 3907 BP.
XX AC ABK34640;
XX DT 08-MAY-2002 (first entry)
XX DE Human cDNA for novel secreted protein, SEQ ID 409.
XX
XX Human; ss; gene; secreted protein; immune deficiency; viral infection;
KW bacterial infection; fungal infection; autoimmune disorder; burn;
KW rheumatoid arthritis; multiple sclerosis; autoimmune thyroiditis;
KW diabetes; allergy; asthma; neurodegenerative disease; Crohn's disease;
KW Alzheimer's disease; Parkinson's disease; liver fibrosis; tumour;
KW coagulation disorder; haemophilia; inflammatory disorder; ulcer;
KW tissue regeneration; wound healing; haematopoiesis; myeloid deficiency;
KW lymphoid cell deficiency.
XX OS Homo sapiens.
XX
XX WO200177290-A2.
XX PN
XX PD 18-OCT-2001.
XX
XX 29-MAR-2001; 2001WO-US010295.
XX
XX 06-APR-2000; 2000US-0194941P.
XX
XX (GEMY) GENETICS INST INC.
XX
XX Wong GG, Clark HF, Fechtel K, Agostino MJ, Howes SH, Resnick RJ;
PI Gulukota K, Graham JR;
XX
XX WPI; 2002-179323/23.
XX
XX Six hundred and twenty five polynucleotides derived from a variety of
PT human tissue sources which encode secreted proteins, useful for treating
PT immune deficiencies and disorders such as autoimmune disorders.
XX
XX Claim 1; Page 210-211; 339pp; English.

XX The invention relates to 625 polynucleotides which have been derived from
 CC a variety of human tissue sources and which encode novel secreted
 CC proteins, their complements and sequences that hybridize to them. Also
 CC included are a vector comprising the polynucleotide, a host cell
 CC transformed with the vector, the proteins encoded by the polynucleotides,
 CC antibodies that bind to the proteins and identification of modulators of
 CC the proteins or the expression of the polynucleotide. The polynucleotides
 CC can be used as probes for the identification and isolation of full length
 CC cDNA and genomic DNA. The polynucleotides and proteins can also be used
 CC as nutritional supplements. The protein is useful in the treatment of
 CC various immune deficiencies and disorders such as viral infections,
 CC bacterial infections, fungal infections, autoimmune disorders (e.g.
 CC rheumatoid arthritis, multiple sclerosis, autoimmune thyroiditis and
 CC diabetes) and allergic reactions and conditions (e.g. asthma). They are
 CC also useful for treating neurodegenerative diseases (e.g. Alzheimer's
 CC disease, Parkinson's disease), liver fibrosis, coagulation disorders
 CC (e.g. haemophilia), inflammatory disorders (e.g. Crohn's disease) and
 CC tumours. They are also useful for tissue regeneration, for wound healing
 CC and in the treatment of burns, incisions and ulcers. The proteins are
 CC also useful for regulating haematopoiesis, for treating myeloid or
 CC lymphoid cell deficiencies. The present sequence is one of the 625 cDNA
 CC sequences encoding a secreted protein
 XX
 SQ Sequence 3907 BP; 1211 A; 990 C; 820 G; 885 T; 0 U; 1 Other;

Query Match 30.3%; Score 390.4; DB 6; Length 3907;
 Best Local Similarity 62.4%; Pred. No. 2.2e-103;
 Matches 673; Conservative 0; Mismatches 381; Indels 24; Gaps 3;

QY 52 AGCATGGTGGCGCCCTCGCCAAACACAGCCCGCCCTGCGCATCATCAACCATCAGC 111
 DB 2425 AACAACTTGGGATCATCATGACCACTCACCGAGCCCTTCCTCATCAACCGCTGAGC 2484
 QY 112 GCCTGCGAGTCGCTCTGAACAATCACTTGGCAAGCCAGGAGCCCTTGGCGTCACT 171
 DB 2485 GCTTTGAGTCCATCATGACCACTTGGCAAGGTTGTCAGCCCGT----- 2535
 QY 172 TCCTGCTCCAGCCCAAGTTCAAGCAAAATTTCCATGTTTCCCAAGTCGAATCTCAATGTC 231
 DB 2536 -----AGTCCCTCGTGGACCGCTGGCGATGCTGTACAAGATCAGCAACAGCATG 2586
 QY 232 ATGGACAGCCGGTCTTGAGTCTCGCTCCCAAGGTGAGCGGTGTCAGGCGGTAC 291
 DB 2587 CTGGACAAGCCGGTGTACCCGCCCACTTGTGAAGCAGGCGGATGCCATCGACGCTAC 2646
 QY 292 CTGTTTGAAGACAGCGATCAGCCCAATGACCTGACCAAGTCCAAAGCAAGCAAGCCGAG 351
 DB 2647 TATTATGAACAGCGACCGACCCCAATGACTTAACTCAAGTCCAAAGTCCAAAGCCGCTGGTG 2706
 QY 352 TCCTGCGAAGCAATCTTTGATGTCCCACTCAGAAAGCAGGCTCTGTCGACATCGCC 411
 DB 2707 TCCAGGTTGGCTGATTCGGTGGCATCACTCTCGCGGAGAGCGCACTCATGACATCTCC 2766
 QY 412 GACATGTTCAAGTCTCTCCCAAGCCACCCCAAGCCAGCCCTCTCTCTCCAGGCTC 471
 DB 2767 GACATGTTGAAGAACTCAGAGCCGCTGAGCGCCCAAGTCTCCACGCGCTTCCACAGTT 2826
 QY 472 CCCCCCATGAGCTGGAATGATGTCAGGCGCTTTGAGGATGTCCTCCAGTGAAGTCTCA 531
 DB 2827 TCAGA---GAATCGATGCTGATGGCAGACCTTTGAGGAGCGCTTGACGAGCTGTCA 2883
 QY 532 ACTTTGCAATAAGAAAGCCGCGAGTCCCACTGGAATCTCTCAGCATCTTCTGATTTCTA 591
 DB 2884 CCGGTCCACAGAGGAGGCGCGAGTCCCACTGGAACCCGCGAGCACTTCTCATCTCG 2943
 QY 592 CAAGCCAGTTTGGCTTCAGAGCTCTTCCAGACATCAGAGGGCAAAATACCTGCTGTCTGAT 651
 DB 2944 CAGGCCAGTTTGGCTTCAGAGCTTGGCGGAGACCAAGAGGGGCAAGTACATCATGTGCGGAC 3003
 QY 652 CTGGGCCCCAAGAGCGTATGAATCTTAAGTTTACGGACTCTCAATGACCACTATC 711
 DB 3004 TTGGGCCCCGAGGAGGAGGTGCATCTCGAAGTTTACTGGGCTCTCCATGACCACTC 3063

QY 712 AGTCATGGTGGCGCAACGCTCAAGTACCAAGTCTAGGAAAAACGGCGGAGCAAAATTTCTG 771
 DB 3064 AGCCACTGGCTGGCCAAATGTGAAGTACCAGTTGAGGAGGACAGGCGGAAACCAAAATTCCTA 3123
 QY 772 AAAAACAATGGAAGGAGCCGCCCATCTTTTATTCAGTCTGCTGCTCCAGTTCCAGA 831
 DB 3124 AAGAACCTGGACACAGGCGATCTCTTTTCTTTTGAACAGATTTGCTCTCAGTTCCAGA 3183
 QY 832 ACCCTCTTACCTACATCAGTCACTTAGAATCTCACCTGGGTTTCCAAATGAAGACATG 891
 DB 3184 ACTGCTTCTACATACATAAGTCAITTTGAGACACACACTTGGGCTTCAGCTGAAGATCTC 3243
 QY 892 ACCCGCTTGTGAGTGACGACAAAGCAAGGTGGAGCAAGAGATCTCCCGGTATCGTGG 951
 DB 3244 TCCAAGCTGCCACT---CAATCAGATTCAAGAACAGCAGAAATGTTTCGAAAGTCCCTCACC 3300
 QY 952 GCTCAGAGGTCTCCAGAACCAATAGCTGCCGAGGAGGACACAGACTCTAAATTTCAAGTGT 1011
 DB 3301 AACAAACTCTGGGCGCCACTGGGCGCCACCGAGGAGAGACTTGGGCTCCACATTTCCAATGT 3360
 QY 1012 AAGTTGTGCTGTGCGACATTTGTGAGCAAAACATGCGGTAAAACTCCACTAAGCAAAAGC 1071
 DB 3361 AAGCTCTGCAACCGGACTTTTTCGAGCAAGCAGCAGTCAAACTGCACCTTAGTAAGACC 3420
 QY 1072 CACAGCAAGTCAACCGAACACCAATTCACAGTTTGTAAACAGACGTTGGATGAAGAAATAGC 1129
 DB 3421 CACGCAAGTCTCCCGAGGACCACTGATCTATGTGACTGATGTGAGTTGGAGAAACAGTAGC 3478
 RESULT 14
 ADF55427
 ID ADF55427 standard; DNA; 7081 BP.
 XX
 AC ADF55427;
 XX
 DT 12-FEB-2004 (first entry)
 XX
 DE Human DNA encoding a novel polypeptide #40.
 KW ds; gene; human; brain disease; mental disorder.
 XX
 OS Homo sapiens.
 XX
 PN JP2003245081-A.
 XX
 PD 02-SEP-2003.
 XX
 PF 25-FEB-2002; 2002JP-00047501.
 XX
 PR 25-FEB-2002; 2002JP-00047501.
 XX
 PA (KAZU-) ZH KAZUSA DNA KENKYUSHO.
 XX
 WIPI; 2003-857164/80.
 DR P-PSDB; ADF55476.
 XX
 PT New DNA derived from KG-1 cDNA library and encoded polypeptide, useful for treating mental disorders.
 XX
 PS Claim 1; SEQ ID NO 40; 401pp; Japanese.
 XX
 CC The invention relates to a DNA which encodes a novel polypeptide. A vector containing the DNA is useful as a reagent in estimation of standard substance. The antibody is useful for detecting the vector containing the DNA and for screening substances and compounds that interact specifically with the vector containing the DNA. The vector containing the DNA is used as a pharmaceutical, in treatment or as preventive agent with respect to disease e.g., brain diseases preferably mental disorders. The present sequence represents human DNA encoding a novel polypeptide.
 CC
 SQ Sequence 7081 BP; 1789 A; 1895 C; 1756 G; 1641 T; 0 U; 0 Other;

Query Match				30.3%; Score 390.4; DB 10; Length 7081;			
Best Local Similarity				62.4%; Pred. No. 3e-103;			
Matches 673; Conservative				0; Mismatches 381; Indels 24; Gaps 3;			
QY	52	AGCAATGGGTGCCCTCGCAACACGCGCCCGSCCTGCGATGCATCAACCCACTCAGC	111				
DB	4726	AACAACCTGGGATCATCTGGACCACTCACCGGAGCCTTCTTTCATCAACCCGCTGAGC	4785				
QY	112	GCCTCGAGTCCGTCTGTGAACAATCACTTGGGCAAGCCACGGAGCCCTTGGCGTCACT	171				
DB	4786	GCTTTGAGTCCATCATGAACACCCACTGGCAAGGTGTCCAAGCCCGTG-----	4836				
QY	172	TCCTGTCTCAGCCCAAGTTCAAGCAAAATTTCCATGTTCCACAAGTCGAATCTCAATGTC	231				
DB	4837	-----AGTCCCTCGCTGGACCCCGCTGGCGATGCTGTACAAGATCAGCAACAGCATG	4887				
QY	232	ATGGACAAGCCGCTTTGAGTCTCTGCTCCACAAGGTTCAGCCAGCGTGTCCAGGCGCTAC	291				
DB	4888	CTGGACAAGCCGCTGTACCCCGCACCCCTGTGAAGCAGGCGCGATGCCATCGACCGCTAC	4947				
QY	292	CTGTTTGTAGAACAGCGATCAGCCCATTTGACTGACCAAGTTCAAAGCAAGAAAGCCGAG	351				
DB	4948	TATTATGAAGAACAGCAGCAGCCCATTTGACTTAACCAAGTCCAAAGTCCAAAGCAAGCCGCTG	5007				
QY	352	TCCTCGCAAGCAAACTCTGTATGTCCCACTCAGAAAGCAGCGCTCTGTCTGACATCGCC	411				
DB	5008	TCGACGCTGCTGATTCGGTGGCATCACCTCTCGGGAGAGCGCATCTGAGCATCTCC	5067				
QY	412	GACATGTTCAAAGTCTCTCCCAAGAGCCACCAACCCCAAGAGCGCTCTCTCTCAGGGTCT	471				
DB	5068	GACATGTTGAAAAACCTTCAGAGGCGCGCTGACGCCCAAGTCTCCAGCGCTCCACAGTT	5127				
QY	472	CCCCCATGAGCTGGAATGATGTGAGGCGCTTTCAGAGATGTCTCCAGTGAAGTCTCA	531				
DB	5128	TCAGA---GAAAGTCCGATGCTGATGTCAGCAGCGCTTTGAGGAGGCGTTGGACAGCTGTCA	5184				
QY	532	ACTTTGCATAAAGAAAGGCGCGGAGTCCAACTGGAATCTCTCAGCATCTTCTCATTTCTA	591				
DB	5185	CGGTTCCACAAGAGGAGGCGCGGAGTCCAACTGGAACCCGACAGCACCTTCTCATCTG	5244				
QY	592	CAAGCCGAGTTTGGCTCGAGCCTCTTCAGACATCAGAGGGCAATACCTGCTGTCTGAT	651				
DB	5245	CAGGCCAGTTCCGCTCGAGCTTTCGGGGAGACCAAGAGGGCAAGTACATCATGTCCGAC	5304				
QY	652	CTGGGCCCAAGAGCGGTATGCAAACTCTAAGTTTACGGGACTCTCAATGACCACTATC	711				
DB	5305	TTGGGCCCGAGAGAGGGTGCACATCTCGAAGTTTACTGGGCTCTCCATGACCAACCATC	5364				
QY	712	AGTCACTGGCTGGCCAAAGTCAAGTACCAGCTTAGGAAAAAGCGGGGGAACAAAATTTCTG	771				
DB	5365	AGCCACTGGCTGGCCAAATGTGAATACCAAGTGAAGAGGACAGGGGGAACGAAAATTCCTA	5424				
QY	772	AAAAATGCAAAAGGCCACCCCATCTTTTATTGCAGTGACTGTGCTCCAGTTCCAGA	831				
DB	5425	AAGAACTGGACACAGGGCATCTCGTTTCTTTTGGCAACAGATTGTGCTCTCAGTTTCA	5484				
QY	832	ACCCCTTCTACATCATCACTGACTTAGAATCTCACCTGGGTTTCCAAATGAAGGACATG	891				
DB	5485	ACTGCTTCTACATACATAAGTCAATTTGGAGACACACTTTGGGCTTCAGCCTGAAGGATCTC	5544				
QY	892	ACCCGCTTGTGAGTGGACCAAGCAAGGTGGAGCAAGATCTCTCCCGGTTATCGTCG	951				
DB	5545	TCCAAGCTGCCACT---CAATTCAGATTCAGAACAGCAGAGATGTTTGGAAAGTCTCTCACC	5601				
QY	952	GCTCAGAGGTCTCCAGAAAAAATAGCTGCCGAGAGGACACAGACTCTAAATTTCAAGTGT	1011				
DB	5602	AACAAAATCTGGGCCCACTGGGGGCCACCGAGGAAGACTTGGGCTCCACATTTCCAATGT	5661				
QY	1012	AAGTTGTGCTGCGACATTTTGTGAGCAACATGCGGTAAACTCCACTTAAGCAAAAGC	1071				
DB	5662	AAGCTCTGCAACCGGACTTTTGGAGCAAGCAGCAGTCAAACTGACCTTAGTAAGACC	5721				

QY	1072	CACAGCAAGTCCACCGAACACCACTTACAGTTTGTAAACAGACGTGATGAAGAATAGC	1129
DB	5722	CAGGCAAGTCTCCCGAGGACCACTGATCTATGTGACTGTGAGAAACAGTAGC	5779
RESULT 15			
ID	ADC31952		
XX	ADC31952 standard; cDNA; 1079 BP.		
AC	ADC31952;		
DT	18-DEC-2003 (first entry)		
XX	Human novel cDNA contig sequence, SEQ ID NO:2034.		
DE	Human; diagnostic; drug screening; forensics; gene mapping;		
KW	biodiversity assessment; Parkinson's disease; Alzheimer's disease;		
KW	neurodegenerative diseases; anaemia; platelet disorder; wound; burns;		
KW	ulcers; osteoporosis; autoimmune disease; cancer;		
KW	molecular weight marker; food supplement; antiparkinsonian; nootropic;		
KW	neuroprotective; anti-anaemic; anticoagulant; thrombolytic; vulnery;		
KW	antiulcer; osteopathic; immunosuppressive; antiinflammatory; cytostatic;		
KW	gene therapy; chromosome 20; ss.		
OS	Homo sapiens.		
XX	WO2003029271-A2.		
PN	10-APR-2003.		
XX	24-SEP-2002; 2002WO-US030474.		
PF	24-SEP-2001; 2001US-0324631P.		
PR	(HYSE-) HYSEQ INC.		
XX	Tang TY, Zhang J, Ren P, Xue AJ, Zhao QA, Wang J, Wehrman T;		
PI	Zhou P, Ghosh M, Wang D, Ma Y, Asundi V, Wang Z, Weng G;		
PI	Haley-Vicente D, Drmanac RT;		
XX	WPI; 2003-371981/35.		
DR	P-PSDB; ADC32719.		
XX	New polynucleotide and polypeptide useful for diagnosing, preventing or		
PT	treating conditions such as neurodegenerative diseases, anemias, platelet		
PT	disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or		
PT	cancer.		
XX	Example 2; SEQ ID NO 2034; 1185pp; English.		
PS	The invention relates to 971 novel human cDNA sequences (ADC29919-		
XX	ADC30889) and the polypeptides they encode (ADC30890-ADC31860). The		
CC	invention also relates to nucleic acid sequences over 99% identical with		
CC	the novel human cDNAs. The invention additionally encompasses expression		
CC	vectors and host cells comprising a nucleic acid of the invention; the		
CC	recombinant production of a polypeptide of the invention; an antibody		
CC	against a polypeptide of the invention; a method of detecting		
CC	polynucleotides or polypeptides of the invention; and methods of		
CC	identifying a compound which binds to a polypeptide of the invention. The		
CC	invention further discloses methods of preventing, treating or		
CC	ameliorating a medical condition; kits comprising polynucleotide probes		
CC	and/or monoclonal antibodies for carrying out the methods of the		
CC	invention; methods for the identification of compounds that modulate the		
CC	expression or activity of the polynucleotide and/or polypeptide; and 767		
CC	contig sequences corresponding to the cDNA sequences of the invention		
CC	(ADC31861-ADC32627) and the polypeptides encoded by the contigs (ADC32628		
CC	-ADC33394). The nucleic acids and polypeptides of the invention are		
CC	useful in diagnostics, drug screening, forensics, gene mapping, in the		
CC	identification of mutations responsible for genetic disorders or other		
CC	traits, for assessing biodiversity, and in producing many other types of		
CC	data and products dependent on DNA and amino acid sequences. They are		
CC	also used for treating diseases such as Parkinson's disease, Alzheimer's		
CC	disease and other neurodegenerative diseases, anaemia, platelet		

CC disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or
CC cancer. The nucleic acids may also be used as hybridisation probes or
CC primers, and in the recombinant production of a protein. The polypeptides
CC are also useful in generating antibodies, as molecular weight markers,
CC and as food supplements. The present sequence represents a human contig
CC sequence used in an example of the invention. Note: The sequence data for
CC this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.

XX

SQ Sequence 1079 BP; 322 A; 317 C; 257 G; 183 T; 0 U; 0 Other;

Query Match 30.3%; Score 390; DB 10; Length 1079;
Best Local Similarity 100.0%; Pred. No. 1.5e-103;
Matches 390; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAGGCGCAGCGAGAGAGAAACCCAGCCCTGGAGCCCAACATCTGCTCTGAGCAATGGG 60
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
617 GAGGCGCAGCGAGAGAGAAACCCAGCCCTGGAGCCCAACATCTGCTCTGAGCAATGGG 676
QY 61 TCGGCCCTCGCCAAACAGCCCGCCCTGCGCATCATCAACCCACTCAGCGCCCTGCAG 120
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
677 TCGGCCCTCGCCAAACAGCCCGCCCTGCGCATCATCAACCCACTCAGCGCCCTGCAG 736
QY 121 TCCGTCCTGAACAATCACTTGGGCAAGCCAGGAGCCCTTGGCTCAGCTTCCTGCTCC 180
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
737 TCCGTCCTGAACAATCACTTGGGCAAGCCAGGAGCCCTTGGCTCAGCTTCCTGCTCC 796
QY 181 AGCCCAAGTTCAAGCACAAATTTCCATGTTCCACAAGTCGAATCTCAATGTCAATGACAAG 240
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
797 AGCCCAAGTTCAAGCACAAATTTCCATGTTCCACAAGTCGAATCTCAATGTCAATGACAAG 856
QY 241 CCGGTCTTGAAGTCTGCTCCACAAAGGTCAGCCAGCGGTGTCAGGCGCTACCTGTTTGAG 300
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
857 CCGGTCTTGAAGTCTGCTCCACAAAGGTCAGCCAGCGGTGTCAGGCGCTACCTGTTTGAG 916
QY 301 AACAGGGATCAGCCCATTTGACCTGACCAAGTCCAAAAGCAAGAGCCGAGTCTCGCAA 360
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
917 AACAGGGATCAGCCCATTTGACCTGACCAAGTCCAAAAGCAAGAGCCGAGTCTCGCAA 976
QY 361 GCACAATCTTGATGTGCCCACTCAGAAG 390
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
977 GCACAATCTTGATGTGCCCACTCAGAAG 1006

Search completed: May 5, 2005, 00:37:17
Job time : 901.572 secs

This Page Blank (uspto)

GenCore version 5.1.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 4, 2005, 23:46:20 ; Search time 289.356 Seconds
(without alignments)
7283.499 Million cell updates/sec

Title: US-08-731-499-5
Perfect score: 1288
Sequence: 1 GAGGCACCGAGAGAGAA.....AAAAAAAAAAAAAAAAATTC 1288

Scoring table: IDENTITY_NUC
Gapop 10_0 , Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA.*
1: /cgn2_6/ptodata/1/ina/5A COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PCTUS COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1288	100.0	1288	2	US-08-680-395-5
2	1288	100.0	1288	4	US-08-892-695-5
3	377.8	29.3	2918	4	US-09-510-543-21
4	256.4	19.9	1144	2	US-08-828-008-1
5	113	8.8	282	2	US-08-828-008-3
6	41.6	3.2	258	4	US-09-023-655-773
7	40	3.1	31623	4	US-09-949-016-15945
8	37.6	2.9	601	4	US-09-949-016-115533
9	37.6	2.9	601	4	US-09-949-016-115626
10	37.6	2.9	601	4	US-09-949-016-115719
11	37.6	2.9	601	4	US-09-949-016-115812
12	37.6	2.9	601	4	US-09-949-016-115905
13	37.6	2.9	601	4	US-09-949-016-115998
14	37.6	2.9	601	4	US-09-949-016-116091
15	37.6	2.9	601	4	US-09-949-016-116184
16	37.6	2.9	601	4	US-09-949-016-116277
17	37.6	2.9	601	4	US-09-949-016-116370
18	37.6	2.9	110266	4	US-09-949-016-14913
19	37.6	2.9	110266	4	US-09-949-016-14914
20	37.6	2.9	110266	4	US-09-949-016-14915
21	37.6	2.9	110266	4	US-09-949-016-14916
22	37.6	2.9	110266	4	US-09-949-016-14917
23	37.6	2.9	110266	4	US-09-949-016-14918
24	37.6	2.9	110266	4	US-09-949-016-14919
25	37.6	2.9	110266	4	US-09-949-016-14920
26	37.6	2.9	110266	4	US-09-949-016-14921
27	37.6	2.9	110266	4	US-09-949-016-14922

c 28	37	2.9	505	4	US-09-621-976-15639	Sequence 15639, A
c 29	37	2.9	1389	4	US-09-902-540-3663	Sequence 3663, Ap
c 30	37	2.9	23847	4	US-09-902-540-1177	Sequence 1177, Ap
c 31	36.8	2.9	118136	4	US-09-949-016-12439	Sequence 12439, A
c 32	36.2	2.8	399	4	US-09-621-976-8976	Sequence 8976, Ap
c 33	36.2	2.8	1497	4	US-09-220-132-94	Sequence 94, Appl
c 34	35.8	2.8	1905	4	US-09-902-540-2711	Sequence 2711, Ap
c 35	35.8	2.8	14861	4	US-09-902-540-1127	Sequence 1127, Ap
c 36	35.4	2.7	1141	4	US-09-889-463A-37	Sequence 37, Appl
c 37	35.4	2.7	7007	4	US-09-949-016-12430	Sequence 12430, A
c 38	35.4	2.7	7007	4	US-09-949-016-15989	Sequence 15989, A
c 39	35.4	2.7	50836	4	US-09-949-016-16722	Sequence 16722, A
c 40	35.2	2.7	2406	4	US-09-902-540-5545	Sequence 5545, Ap
c 41	35.2	2.7	11812	4	US-09-902-540-1041	Sequence 1041, Ap
c 42	35.2	2.7	26452	4	US-09-949-016-15822	Sequence 15822, A
c 43	35.2	2.7	64190	4	US-09-949-016-14712	Sequence 14712, A
c 44	35.2	2.7	64190	4	US-09-949-016-14713	Sequence 14713, A
c 45	35.2	2.7	133613	4	US-09-949-016-15824	Sequence 15824, A

ALIGNMENTS

RESULT 1
US-08-680-395-5
; Sequence 5, Application US/08680395
; Patent No. 5892010
; GENERAL INFORMATION:
; APPLICANT: Gray, Joe W.
; APPLICANT: Collins, Colin
; APPLICANT: Hwang, Soo-in
; APPLICANT: Godfrey, Tony
; APPLICANT: Kowbel, David
; APPLICANT: Rommens, Johanna
; TITLE OF INVENTION: Genes from the 20q13 Amplicon and Their
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICANT NUMBER: US/08/680.395
FILING DATE: 15-JUL-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Bastian, Kevin L.
REGISTRATION NUMBER: 34,774
REFERENCE/DOCKET NUMBER: 023070-06890005
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1288 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: 1..1288 /note= "cDNA clone 41.1 with homology
OTHER INFORMATION: to homeobox T shirt gene from
OTHER INFORMATION: Drosophila"

US-08-680-395-5									
Query Match 100.0%; Score 1288; DB 2; Length 1288;									
Best Local Similarity 100.0%; Pred. No. 0;									
Matches 1288; Conservative 0; Mismatches 0; Indels 0; Gaps 0;									
QY	1	GAGGCGAGCAGAGGAGAAACCCAGCCCTGGAGCCACATCTCTCTGAGCAATGGG	60						
DB	1	GAGGCGAGCAGAGGAGAAACCCAGCCCTGGAGCCACATCTCTCTGAGCAATGGG	60						
QY	61	TGGCCCTCGCCAAACACGCGCCCGGCTGGCCATGATCAACCCACTAGCGCCCTGCAG	120						
DB	61	TGGCCCTCGCCAAACACGCGCCCGGCTGGCCATGATCAACCCACTAGCGCCCTGCAG	120						
QY	121	TCGGTCTGAAACATCACTTGGGCAAGCCAGGAGCCCTTGGCTCACTCTCTGCTCC	180						
DB	121	TCGGTCTGAAACATCACTTGGGCAAGCCAGGAGCCCTTGGCTCACTCTCTGCTCC	180						
QY	181	AGCCCAAGTTCAAGCAAAATTTCCATGTTTCCACAAGTCGAATCTCAATGTATGACAAG	240						
DB	181	AGCCCAAGTTCAAGCAAAATTTCCATGTTTCCACAAGTCGAATCTCAATGTATGACAAG	240						
QY	241	CGGTCCTTGAGTCTCGCTCCACAAGGTACGCGGTGTCAGGCGCTACCTGTTTGA	300						
DB	241	CGGTCCTTGAGTCTCGCTCCACAAGGTGTCAGGCGGTGTCAGGCGCTACCTGTTTGA	300						
QY	301	AACAGCGATCAGCCCATTGACCTGACCAAGTCCAAAGCAAGAAAGCCAGTCTCGCAA	360						
DB	301	AACAGCGATCAGCCCATTGACCTGACCAAGTCCAAAGCAAGAAAGCCAGTCTCGCAA	360						
QY	361	GCAAAATCTTGATGTGTCCTCCACCTCAGAAAGCAGCTCTGTCTGACATCGCCGACATGGTC	420						
DB	361	GCAAAATCTTGATGTGTCCTCCACCTCAGAAAGCAGCTCTGTCTGACATCGCCGACATGGTC	420						
QY	421	AAAGTCTCTCCAAAGCCACCCCAAGCCAGCTCTCTCTCCAGGGTCCCGCCATG	480						
DB	421	AAAGTCTCTCCAAAGCCACCCCAAGCCAGCTCTCTCTCCAGGGTCCCGCCATG	480						
QY	481	AAGCTGGAATGATGTGAGGCGCTTTGAGGATGTCTCCAGTGAAGTCTCACTTTGCAT	540						
DB	481	AAGCTGGAATGATGTGAGGCGCTTTGAGGATGTCTCCAGTGAAGTCTCACTTTGCAT	540						
QY	541	AAAAGAAAAGCGCGCAGTCCAACTGGAATCTCTCAGCATCTCTGATCTTACAAAGCCAG	600						
DB	541	AAAAGAAAAGCGCGCAGTCCAACTGGAATCTCTCAGCATCTCTGATCTTACAAAGCCAG	600						
QY	601	TTTGCTCTGAGCTCTTCCAGCATCAGAGGGCAATATCTGCTGTCTGATCTGGGCCCA	660						
DB	601	TTTGCTCTGAGCTCTTCCAGCATCAGAGGGCAATATCTGCTGTCTGATCTGGGCCCA	660						
QY	661	CAAGCGGTATGAAATCTTAAGTTTACGGGACTCTCAATGACCACTATCAGTCACTGG	720						
DB	661	CAAGCGGTATGAAATCTTAAGTTTACGGGACTCTCAATGACCACTATCAGTCACTGG	720						
QY	721	CTGGCCAAACGTCAAGTACAGCTTAGGAAAACGGGGGACAAAATTTCTGAAAAACATG	780						
DB	721	CTGGCCAAACGTCAAGTACAGCTTAGGAAAACGGGGGACAAAATTTCTGAAAAACATG	780						
QY	781	GACAAAGGCCACCCCATCTTTTATGAGTGAATGTGCTCTCCAGTTCAAGACCCCTTCT	840						
DB	781	GACAAAGGCCACCCCATCTTTTATGAGTGAATGTGCTCTCCAGTTCAAGACCCCTTCT	840						
QY	841	ACCTACATCAGTCACTTAGAATCTCACCTGGGTTTCCAAATGAAGACATGACCCGCTTG	900						
DB	841	ACCTACATCAGTCACTTAGAATCTCACCTGGGTTTCCAAATGAAGACATGACCCGCTTG	900						
QY	901	TCAGTGGACCAAGCAAGGTGGAGCAAGATCTCCCGGTATCGTGGCTCAGAGG	960						
DB	901	TCAGTGGACCAAGCAAGGTGGAGCAAGATCTCCCGGTATCGTGGCTCAGAGG	960						
QY	961	TCTCCAGAAAACAATAGTCCGGAAGGACACAGACTCTAAATTCAGTGTAAAGTTGTGC	1020						
DB	961	TCTCCAGAAAACAATAGTCCGGAAGGACACAGACTCTAAATTCAGTGTAAAGTTGTGC	1020						

US-08-892-695-5									
Query Match 100.0%; Score 1288; DB 4; Length 1288;									
Best Local Similarity 100.0%; Pred. No. 0;									
Matches 1288; Conservative 0; Mismatches 0; Indels 0; Gaps 0;									
QY	1	GAGGCGAGCAGAGGAGAAACCCAGCCCTGGAGCCACATCTCTCTGAGCAATGGG	60						
DB	1	GAGGCGAGCAGAGGAGAAACCCAGCCCTGGAGCCACATCTCTCTGAGCAATGGG	60						
QY	61	TGGCCCTCGCCAAACACGCGCCCGGCTGGCCATGATCAACCCACTAGCGCCCTGCAG	120						
DB	61	TGGCCCTCGCCAAACACGCGCCCGGCTGGCCATGATCAACCCACTAGCGCCCTGCAG	120						
QY	121	TCGGTCTGAAACATCACTTGGGCAAGCCAGGAGCCCTTGGCTCACTCTCTGCTCC	180						
DB	121	TCGGTCTGAAACATCACTTGGGCAAGCCAGGAGCCCTTGGCTCACTCTCTGCTCC	180						
QY	181	AGCCCAAGTTCAAGCAAAATTTCCATGTTTCCACAAGTCGAATCTCAATGTATGACAAG	240						
DB	181	AGCCCAAGTTCAAGCAAAATTTCCATGTTTCCACAAGTCGAATCTCAATGTATGACAAG	240						
RESULT 2									
US-08-892-695-5									
Sequence 5, Application US/08892695A									
Patent No. 6808878									
GENERAL INFORMATION:									
APPLICANT: Gray, Joe W									
APPLICANT: Collins, Collin									
APPLICANT: Hwang, Soo In									
APPLICANT: Godfrey, Tony									
APPLICANT: Kowel, David									
APPLICANT: Rommens, Johanna									
TITLE OF INVENTION: GENES FROM THE 20Q13 AMPLICON AND THEIR USES									
FILE REFERENCE: 2500.124US3									
CURRENT APPLICATION NUMBER: US/08/892,695A									
CURRENT FILING DATE: 1997-07-15									
EARLIER APPLICATION NUMBER: 08/785,532									
EARLIER FILING DATE: 1997-01-17									
EARLIER APPLICATION NUMBER: 08/731,499									
EARLIER FILING DATE: 1996-10-16									
EARLIER APPLICATION NUMBER: 08/680,395									
EARLIER FILING DATE: 1996-07-15									
NUMBER OF SEQ ID NOS: 59									
SOFTWARE: PatentIn Ver. 2.0									
SEQ ID NO 5									
LENGTH: 1288									
TYPE: DNA									
ORGANISM: Artificial Sequence									
FEATURE:									
OTHER INFORMATION: Description of Artificial Sequence:41.1									
US-08-892-695-5									
Query Match 100.0%; Score 1288; DB 4; Length 1288;									
Best Local Similarity 100.0%; Pred. No. 0;									
Matches 1288; Conservative 0; Mismatches 0; Indels 0; Gaps 0;									
QY	1	GAGGCGAGCAGAGGAGAAACCCAGCCCTGGAGCCACATCTCTCTGAGCAATGGG	60						
DB	1	GAGGCGAGCAGAGGAGAAACCCAGCCCTGGAGCCACATCTCTCTGAGCAATGGG	60						
QY	61	TGGCCCTCGCCAAACACGCGCCCGGCTGGCCATGATCAACCCACTAGCGCCCTGCAG	120						
DB	61	TGGCCCTCGCCAAACACGCGCCCGGCTGGCCATGATCAACCCACTAGCGCCCTGCAG	120						
QY	121	TCGGTCTGAAACATCACTTGGGCAAGCCAGGAGCCCTTGGCTCACTCTCTGCTCC	180						
DB	121	TCGGTCTGAAACATCACTTGGGCAAGCCAGGAGCCCTTGGCTCACTCTCTGCTCC	180						
QY	181	AGCCCAAGTTCAAGCAAAATTTCCATGTTTCCACAAGTCGAATCTCAATGTATGACAAG	240						
DB	181	AGCCCAAGTTCAAGCAAAATTTCCATGTTTCCACAAGTCGAATCTCAATGTATGACAAG	240						

```
QY 241 CCGGCTCTTGAGTCCCTGCTCCACAGGTCAGCCAGGCTCCAGGGCTACTCTGTTGAG 300
Db 241 CCGGCTCTTGAGTCCCTGCTCCACAGGTCAGCCAGGTCAGCCAGGCTACTCTGTTGAG 300
QY 301 AACAGCGATCAGCCCAATGACCTGACCAAGTCCAAAAGCAAGAAAGCCGAGTCCTCGCAA 360
Db 301 AACAGGGATCAGCCCAATGACCTGACCAAGTCCAAAAGCAAGAAAGCCGAGTCCTCGCAA 360
QY 361 GCACAATCTTGATGTCCTCCACCTCAGAGCAAGCTCTGTCGATCGCGACATGTCGTC 420
Db 361 GCACAATCTTGATGTCCTCCACCTCAGAGCAAGCTCTGTCGATCGCGACATGTCGTC 420
QY 421 AAAGTCCTCCCAAGCAGCAGCCCAAGCAGCAGCTCTCTCCAGGTCCTCCCGCCATG 480
Db 421 AAAGTCCTCCCAAGCAGCAGCCCAAGCAGCAGCTCTCTCCAGGTCCTCCCGCCATG 480
QY 481 AAGCTGGAAATGGATGTGAGGGCTTTGAGGATGTCTCAGTGAAGTCTCAACTTTGCAT 540
Db 481 AAGCTGGAAATGGATGTGAGGGCTTTGAGGATGTCTCAGTGAAGTCTCAACTTTGCAT 540
QY 541 AAAGAAAGGCGGAGTCCTCAACTGGAATCTCAGCATCTTCTGATCTCAAGCCCGAG 600
Db 541 AAAGAAAGGCGGAGTCCTCAACTGGAATCTCAGCATCTTCTGATCTCAAGCCCGAG 600
QY 601 TTTGGCTCGAGCTCTTCCAGACATCAGAGGCAATACCTGCTCTGATCTGGGCCA 660
Db 601 TTTGGCTCGAGCTCTTCCAGACATCAGAGGCAATACCTGCTCTGATCTGGGCCA 660
QY 661 CAAGAGCGTATGCAATCTCTAAGTTTACGGGACTCTCAATGACCACTATCAGTCACTGG 720
Db 661 CAAGAGCGTATGCAATCTCTAAGTTTACGGGACTCTCAATGACCACTATCAGTCACTGG 720
QY 721 CTGGCCAAACGTCAGTACAGCTTAGGAAACCGGCGGACAAAATTTCTGAAAACATG 780
Db 721 CTGGCCAAACGTCAGTACAGCTTAGGAAACCGGCGGACAAAATTTCTGAAAACATG 780
QY 781 GACAAAGGCCACCCATCTTTTATGTCAGTGTCTGCTCCAGTTCAGAACCCCTTCT 840
Db 781 GACAAAGGCCACCCATCTTTTATGTCAGTGTCTGCTCCAGTTCAGAACCCCTTCT 840
QY 841 ACCTACATCAGTCAGTCTAGAAATCTCACTGGGTTTCCAAATGAAGGACATGACCGCTTG 900
Db 841 ACCTACATCAGTCAGTCTAGAAATCTCACTGGGTTTCCAAATGAAGGACATGACCGCTTG 900
QY 901 TCAGTGGACCAAGCAAGGTCGAGCAAGAGATCTCCCGGATCTGTCGGCTCAGAGG 960
Db 901 TCAGTGGACCAAGCAAGGTCGAGCAAGAGATCTCCCGGATCTGTCGGCTCAGAGG 960
QY 961 TCTCAGAAACATAGCTGCGAAGAGGACAGACTCTAAATTCAGTGTGAAGTTGTGC 1020
Db 961 TCTCAGAAACATAGCTGCGAAGAGGACAGACTCTAAATTCAGTGTGAAGTTGTGC 1020
QY 1021 TGTCCGACATTTGTGAGCAAAACATCGGTAAACCTCCACCTAAGCAAAACGACAGCAAG 1080
Db 1021 TGTCCGACATTTGTGAGCAAAACATCGGTAAACCTCCACCTAAGCAAAACGACAGCAAG 1080
QY 1081 TCACCCGAACACCATTCACAGTTTGTAAACAGAGTGGATGAAGATAGCTCTGAGGAGC 1140
Db 1081 TCACCCGAACACCATTCACAGTTTGTAAACAGAGTGGATGAAGATAGCTCTGAGGAGC 1140
QY 1141 AATGCTTAGTTTCCATTTTCAGGCTGGATCCCTCAGACTGAAACCTTCTGCTGTTGA 1200
Db 1141 AATGCTTAGTTTCCATTTTCAGGCTGGATCCCTCAGACTGAAACCTTCTGCTGTTGA 1200
QY 1201 CCATCTGCTCTGACATTTGAACTCATTTGAACTCTCTGACACCCCTGGCTCTGAGAGA 1260
Db 1201 CCATCTGCTCTGACATTTGAACTCATTTGAACTCTCTGACACCCCTGGCTCTGAGAGA 1260
QY 1261 CTGCCAAAAAATTTTCAATTTTCAATTTTCAATTTTCAATTTTCAATTTTCAATTTTCA 1288
Db 1261 CTGCCAAAAAATTTTCAATTTTCAATTTTCAATTTTCAATTTTCAATTTTCAATTTTCA 1288
```

RESULT 3

US-09-510-543-21
; Sequence 21, Application US/09510543
; Patent No. 6517837

GENERAL INFORMATION:

; APPLICANT: Scanlan, Matthew J.
; APPLICANT: Chen, Yao-Tsang
; APPLICANT: Stockert, Elisabeth

; APPLICANT: Old, Lloyd J.

; TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULES ASSOCIATED WITH COLON CANCER AND

; TITLE OF INVENTION: METHODS FOR DIAGNOSING AND TREATING COLON CANCER

; FILE REFERENCE: LUD-5506.1-JEL/NDH

; CURRENT APPLICATION NUMBER: US/09/510,543

; CURRENT FILING DATE: 2000-02-22

; PRIOR APPLICATION NUMBER: US 09/102,322

; NUMBER OF SEQ ID NOS: 22

; SEQ ID NO 21

; LENGTH: 2918

; TYPE: DNA

; ORGANISM: Homo sapiens

US-09-510-543-21

Query Match 29.3%; Score 377.8; DB 4; Length 2918;

Best Local Similarity 62.3%; Pred. No. 1.8e-109;

Matches 672; Conservative 0; Mismatches 382; Indels 25; Gaps 4;

```
QY 52 AGCAATGGGTGGCGCTCTGCCAAGCCACGCGCCCTGCCATGTCATCAACCCACTCAGC 111
Db 1429 AACAACTCTGGGATCATCATGACCACTCACCAGGAGCTTCTTCTCAACCCGCTGAGC 1488
QY 112 GCCCTGAGTCGCTCTGAAACATCATTTGGGCAAGCCACGGAGCCCTTGGCTCACT 171
Db 1489 GCTTGGCATTCATCATGAACACCCACCTGGGCAAGGTGTCCAAGCCCGTG----- 1539
QY 172 TCCTGCTCCAGCCCAAGTTCAAGCACAAATTTCCATGTTTCCACAAGTCGAATCTCAATGTC 231
Db 1540 -----AGTCCCTCGCTGGACCCGCTGGGATGCTGTACAGATCAGCAACAGCATG 1590
QY 232 ATGGACAAGCCGGTCTTGAGTCTGCTCCCAAGGTCAGCCAGCGGTGTCTCAGGGCGTAC 291
Db 1591 CTGGACAAGCCGGTGTATCCCGCCACCCCTGTGAAGAGCGGCGATGCCATCGACCGCTAC 1650
QY 292 CTGTTTGAAGACAGCGATCAGCCCATTTGACCTGACCAAGTCCAAAAGCAGAAAGCCGAG 351
Db 1651 TATTATGAAAACAGCGACCGACCCATTGACTTAACTCAAGTCCAAAGAACAAAGCCGCTGTG 1710
QY 352 TCCTCGCAAGCACAAATCTTGTATGTCCTCCACCTCAGAGACAGCTCTGTCTGACATCGCC 411
Db 1711 TCCAGCGTGGCTGATTCGGTGGCATCACCTCTGCGGGAGAGCGCACTCATGACATCTCC 1770
QY 412 GACATGGTCAAAAGTCTCTCCCAAGCCACCCCAAGCCAGCGCTCTCTTCCAGGGTTC 471
Db 1771 GACATGGTGAATAAATCTCAGAGCGCTGAGCGCCCAAGTCTCTCAGCCCTCCACAGTT 1830
QY 472 CCCCCATGAAGCTGGAATGATGTGAGGCGCTTTTGGAGTGTCTCCAGTGAAGTCTCA 531
Db 1831 TCAGA---GAAGTCCGATGCTGATGGCAGCAGCTTTGAGGAGCGCTTGGACAGCTGTCA 1887
QY 532 ACTTGTCAATAAAGAAAGGCGGAGTCCAACTGGAATCTCTCAGCATCTCTGATTCTA 591
Db 1888 CCGGTCCAAGAGAGAGGCGGCGAGTCCAACTGGAACCCCGCAGCACCTTCTCATCTCTG 1947
QY 592 CAAGCCCATGTTGCTCGAGCTCTTCCAGACATCAGAGGGCAAAATACCTGCTGTCTGAT 651
Db 1948 CAGGCCAGTTCGCTCTGAGCTTGGCGGAGACCAAGAGGCAAGTACATCATGTCGGAC 2007
QY 652 CTGGGCCCAAGAGCGTATGCAAACTCTTAAGTTTACGGGACTCTCAATGACCACTATC 711
Db 2008 TTGGCCCGCAGGAGAGGGTGCACATCTCGAAGTTTACTGGGCTCTCCATGACCACTATC 2067
QY 712 AGTCATCTGGTGGCCAAAGTCAAGTACAGCTTACGAAAACGGGCGGACAAAATTTCTG 771
```

Db 2068 AGCCACTGGCTGCCCAATGTGAAGTACAGATTGAGGAGGACAGGGGAAACGAAATTCCTA 2127
Qy 772 AAAACATGACAAAGGCCACCCATCTTTTATTGCAGTGACTGTGCTCCCGATTTCAGA 831
Db 2128 AGAAGCTGGACACAGGGGATCTGTTTCTTTTGAAGATTGTGCTCTCAGTTTCCA 2187
Qy 832 ACCCTTTACCTTACATCAGTCACTTGTAGAAATCTCACCTGGGTTTCCAAATGAAGACATG 891
Db 2188 ACTGCTTCTACATACATAGTCAATTTGGAGACACACTTGGCTTACGCTGAAGATCTC 2247
Qy 892 ACCGGTTGTGAGTGGACGAGCAAGGAGGAGCAAGATCTCCCGGGTATCGTG 951
Db 2248 TCAAAGCTGCCACT---CAATCAGATTCAAGAACAGCAGAAATGTTTCGAAAGTCTCTCACC 2304
Qy 952 GCTCAGAGTCTCCAGAAACATAGTCTGCCGAGAGGACACAGACTCTAAATTCAGTGT 1011
Db 2305 ACAAAACTCTTGGGCCCATCTGGGGCCACCGAGGAAGACTTGGGCTCCACATTCCAATGT 2364
Qy 1012 AAGTTGTGCTGGGACATTTGTG-AGCAAAACATGCGGTAAAACTCCACTAAGCAAAAC 1070
Db 2365 AAGCTCTGCAACCGGACTTTTGGAGAGCAAGCAGCAGTCAAACTGCACCTTAGTAAGAC 2424
Qy 1071 GCACGCAAGTCAACCGCAACACCAITTCACAGTTTGTAAACAGAGTGGATGAAGATAGC 1129
Db 2425 CCACGGCAAGTCTCCCGAGGACCACCTGATCTATGTGACTGTGAGTTGGAGAAACAGTAGC 2483

RESULT 4

US-08-828-008-1
; Sequence 1, Application US/08828008
; Patent No. 5939316
; GENERAL INFORMATION:
; APPLICANT: Chen, Hong
; APPLICANT: Friemer, Nelson, B.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR
; TITLE OF INVENTION: THE DIAGNOSIS AND TREATMENT OF NEUROPSYCHIATRIC
; TITLE OF INVENTION: DISORDERS
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036/2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/828,008
; FILING DATE: 27-MAR-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7853-071
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-790-9090
; TELEFAX: 212-869-8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1144 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; NAME/KEY: Coding Sequence

; LOCATION: 1...282
; OTHER INFORMATION:
; NAME/KEY: Fsh 22
; LOCATION: 1...1144
; OTHER INFORMATION:
; US-08-828-008-1

Query Match 19.9%; Score 256.4; DB 2; Length 1144;
Best Local Similarity 67.2%; Pred. No. 5.8e-71;
Matches 437; Conservative 0; Mismatches 206; Indels 7; Gaps 5;

Qy 480 GAAGCTGGAATGAGATGTCAGGCGTTTGAGGATGTCTCCAGTGAAGTCTCAACTTTTGA 539
Db 63 GAAGTCCGATGCTGATGGCAGTAGCTTTGAGGAGCGGTGTGACGAGCTGTCAACGGTCCA 122
Qy 540 TAAAGAAAGAGCGCGGAGTCCAACTGGNAATCTCAGCATCTTCTGATTTCTACAAAGCCCA 599
Db 123 CAAGAGGAAGGCGCGGAGTCCAACTGGAAACCGGAGCA-CTTCTCATCTTGAGGCCCA 181
Qy 600 GTTTGCTCGAGCTCTTCCAGACATCAGAGGGCAAAATACCTGCTGTCTGATCTGGGCC 659
Db 182 GTTCGCTCGAGCTTCCGGGAGACACAGAGGGCAAGTACATCATGTCTGCGACTTGGGCC 241
Qy 660 ACAAGAGCGTATGCAAAATCTCTAAGTTTACGGGACTCTCAATGACCACTATCAGTCACTG 719
Db 242 ACAGGAGAGGGTGCACATCTCGAAGTTTACTGGGCTCTCCATGACCAACCATCAGCCACTG 301
Qy 720 GCTGGCCACAGTCAAGTACCAGCTTAGGAAACGGCGGGGACAAAATTTCTGAAAACAT 779
Db 302 GCTGGCCAAATGTGAAGTACCAGTTGAGGAGGACAGGGGGAAAGAAATCTTAAAGAACCT 361
Qy 780 GGACAAAGGCCACCCCATCTTTTATTGCACTGTGCTGCCAGTTCAGAACCCCTTC 839
Db 362 GGACACAGGGCATCTGTTTCTTTTGGCAACGATTTGCTCTCAGTTGAGAACTGCTTC 421
Qy 840 TACCTACATCAGTCACTTGAATCTCACCTGGTTTCCAAATGAAGGACATGACCGCTT 899
Db 422 TACATACATAAGTCACTTTTGGAGACACACTTGGCTTTCAGCTTGAAGGATCTCTCCAAGCT 481
Qy 900 GTCAGTGGACAGCAAGCAAGAGTGGAGCAAGAGATCTCCGGGTATCTCGGCTCAGAG 959
Db 482 GCCACT---CAATCAGATTCAAGAACAGCAGCAATGTTTGAAGTCTCTCACCACAAAC 538
Qy 960 GTCTCCAGAAACAATAGCTGCCGAGAGGACACAGACTCTAAATTCAGTGTAAAGTTGTG 1019
Db 539 TCTGGGCCCACTGGGGGCCACCGAGGAAGACTTGGCTCCACATTCATGTAAGTCTG 598
Qy 1020 CTGTGGACATTTGTGAGCAAAACATGCGGTAAACCTCCACCTTAAGCAAAACGACAGCAA 1079
Db 599 CAACCGGACTTTTG-GAGCAAGCAGCAGTCAAA-TGCACCTTAGTAAGACCCACGCGC-A 655
Qy 1080 GTCACCGCAACACCATTCACAGTTTGTAAACAGCGTGGATGAGATAGC 1129
Db 656 GTCTCCCGAGGACCACCTGATCTATGTGACTGTGAGTTGGAGAAACAGTAGC 705

RESULT 5
US-08-828-008-3
; Sequence 3, Application US/08828008
; Patent No. 5939316
; GENERAL INFORMATION:
; APPLICANT: Chen, Hong
; APPLICANT: Friemer, Nelson, B.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR
; TITLE OF INVENTION: THE DIAGNOSIS AND TREATMENT OF NEUROPSYCHIATRIC
; TITLE OF INVENTION: DISORDERS
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA


```

; ZIP: 10036/2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: Fast-SEQ Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/828,008
; FILING DATE: 27-MAR-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7853-071
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-790-9090
; TELEFAX: 212-869-8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 282 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
; FEATURE:
; NAME/KEY: Other
; LOCATION: 1...282
; OTHER INFORMATION: Open reading frame to
; OTHER INFORMATION: Feb 22
US-08-828-008-3

Query Match      8.8%; Score 113; DB 2; Length 282;
Best Local Similarity 72.4%; Pred. No. 1.4e-25;
Matches 160; Conservative 0; Mismatches 60; Indels 1; Gaps 1;

Qy 480 GAAGCTGGAATGGATGTCAGCGCGTTTGGAGATCTCCAGTGAAGTCTCAACTTTGCA 539
Db 63 GAAGTCCGATGCTGATGCGAGTAGTTTGGAGGCGGTGGACGAGCTGTACCGGTCCA 122

Qy 540 TAAAGAAAGCCGCGAGTCCAACTGGAATCTTCAGCATCTTCTGATTTACAGGCCCA 599
Db 123 CAAGAGGAAGCGCGGAGTCCAACTGGAACCCGAGCA-CTTCTCATCTCGAGGCCCA 181

Qy 600 GTTGCCTCGAGCTCTTCAGACATCAGAGGGGAAATACCTGCTGTGATCTGGGCCC 659
Db 182 GTTCGCTCGAGCTTGGCGGGAGACCACAGAGGGCAAGTACATCATGTCGAGCTTGGGCC 241

Qy 660 ACAAGAGGTATGCAAAATCTTAAGTTTACGGGACTCTCAA 700
Db 242 ACAGGAGAGGTGCAATCTGGAAGTTTACTGGGCTCTCCA 282

RESULT 6
US-09-023-655-773
; Sequence 773, Application US/09023655
; Patent No. 6607879
; GENERAL INFORMATION:
; APPLICANT: Cocks, Benjamin G.
; APPLICANT: Susan G. Stuart
; APPLICANT: Jeffrey J. Seilhaer
; TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF BLOOD CELL GENE
; NUMBER OF SEQUENCES: 1508
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 PORTER DRIVE
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA

```

```

; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/023,655
; FILING DATE: HEREWITH
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Zeller, Karen J.
; REGISTRATION NUMBER: 37,071
; REFERENCE/DOCKET NUMBER: PA-0001 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 855-0555
; TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 773:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 258 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: PGANOT01
; CLONE: 626570
; US-09-023-655-773

Query Match      3.2%; Score 41.6; DB 4; Length 258;
Best Local Similarity 88.0%; Pred. No. 0.0086;
Matches 44; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 GAGGGCAGCGAGAGGAGAAACCCAGCCCTCGAGGCCACCATCTGCTCT 50
Db 207 GAGGGCAGCGAGAGGAGAAACCCAGCCCTCGAGGCCACCATCTTNT 256

RESULT 7
US-09-949-016-15945
; Sequence 15945, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15945
; LENGTH: 31623
; TYPE: DNA
; ORGANISM: Human
; US-09-949-016-15945

Query Match      3.1%; Score 40; DB 4; Length 31623;
Best Local Similarity 54.9%; Pred. No. 0.66;
Matches 79; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

Qy 362 CACAATCTTGTATGTCCCACTCAGAGAGCGCTCTCTGACATCCCGACATGTC 421
Db 28477 CACACCTTAGCATGGCTTCCCGAGGGCCAGGTTCTGGCTCATGTCTCCCATTCGGCC 28536

```



```
; ORGANISM: Human
US-09-949-016-115812

Query Match          2.9%; Score 37.6; DB 4; Length 601;
Best Local Similarity 65.5%; Pred. No. 0.28;
Matches 55; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

QY 1200 ACCATCCTGCTTCTGACATTGAACCTCTCTGACACCTGGCTCTGAGAAG 1259
DB 384 ACCATCCTGGCTAAACCGTGAACCCCTGGCTAAACCGTGAACCCCTGTCTCTACTAAA 325

QY 1260 ACTGCCAAAAAAAAAAAAAAAAAAAAAA 1283
DB 324 AATACAAAAAAAAAAAAAAAAAAAAA 301

RESULT 12
US-09-949-016-115905/c
; Sequence 115905, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/231,498
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 115905
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-115905

Query Match          2.9%; Score 37.6; DB 4; Length 601;
Best Local Similarity 65.5%; Pred. No. 0.28;
Matches 55; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

QY 1200 ACCATCCTGCTTCTGACATTGAACCTCTCTGACACCTGGCTCTGAGAAG 1259
DB 384 ACCATCCTGGCTAAACCGTGAACCCCTGGCTAAACCGTGAACCCCTGTCTCTACTAAA 325

QY 1260 ACTGCCAAAAAAAAAAAAAAAAAAAAAA 1283
DB 324 AATACAAAAAAAAAAAAAAAAAAAAA 301

RESULT 13
US-09-949-016-115998/c
; Sequence 115998, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/231,498
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
```

```
; SEQ ID NO 115998
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-115998

Query Match          2.9%; Score 37.6; DB 4; Length 601;
Best Local Similarity 65.5%; Pred. No. 0.28;
Matches 55; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

QY 1200 ACCATCCTGCTTCTGACATTGAACCTCTCTGACACCTGGCTCTGAGAAG 1259
DB 384 ACCATCCTGGCTAAACCGTGAACCCCTGGCTAAACCGTGAACCCCTGTCTCTACTAAA 325

QY 1260 ACTGCCAAAAAAAAAAAAAAAAAAAAAA 1283
DB 324 AATACAAAAAAAAAAAAAAAAAAAAA 301

RESULT 14
US-09-949-016-116091/c
; Sequence 116091, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/231,498
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 116091
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-116091

Query Match          2.9%; Score 37.6; DB 4; Length 601;
Best Local Similarity 65.5%; Pred. No. 0.28;
Matches 55; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

QY 1200 ACCATCCTGCTTCTGACATTGAACCTCTCTGACACCTGGCTCTGAGAAG 1259
DB 384 ACCATCCTGGCTAAACCGTGAACCCCTGGCTAAACCGTGAACCCCTGTCTCTACTAAA 325

QY 1260 ACTGCCAAAAAAAAAAAAAAAAAAAAAA 1283
DB 324 AATACAAAAAAAAAAAAAAAAAAAAA 301

RESULT 15
US-09-949-016-116184/c
; Sequence 116184, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/237,768
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
```

```
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 116184
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-116184

Query Match      2.9%; Score 37.6; DB 4; Length 601;
Best Local Similarity 65.5%; Pred. No. 0.28; Indels 0; Gaps 0;
Matches 55; Conservative 0; Mismatches 29;

QY 1200 ACCATCCTGCTTCTGACATTGAACCTCATTGAACCTCCTCTGACACCCCTGGCTCTGAGAAG 1259
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 384 ACCATCCTGGCTAAACGGTGAACCCCTGGCTAAACGGTGAACCCCTGCTCTACTAAA 325
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1260 ACTGCCAAAAA AAAAAAAAAA 1283
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 324 AATACAAAAA AAAAAAAAAA 301
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

Search completed: May 5, 2005, 07:03:17
Job time : 293.689 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 5, 2005, 03:59:12 ; Search time 1770.09 Seconds
(without alignments)
4449.254 Million cell updates/sec

Title: US-08-731-499-5

Perfect score: 1288
Sequence: 1 GAGGCGCGAGAGGAGAA.....AAAAAAAAAAAAAAAAATTC 1288

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 5654200 seqs, 3057283753 residues

Total number of hits satisfying chosen parameters: 11308400

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA.*

```
1: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUBCOMB.seq.*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq.*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq.*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq.*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq.*
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq.*
15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq.*
16: /cgn2_6/ptodata/2/pubpna/US10D_PUBCOMB.seq.*
17: /cgn2_6/ptodata/2/pubpna/US10E_PUBCOMB.seq.*
18: /cgn2_6/ptodata/2/pubpna/US10F_PUBCOMB.seq.*
19: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
20: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq.*
21: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
22: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	DB	ID	Description
1	1288	100.0	1288	8	US-08-731-499-5		Sequence 5, Appli
2	1265	98.2	3322	17	US-10-104-047-373		Sequence 373, App
3	773	60.0	2449	9	US-09-764-864-628		Sequence 628, App
4	770.4	59.8	799	9	US-09-764-864-193		Sequence 193, App
5	656.4	51.0	2012	9	US-09-864-761-33203		Sequence 33203, A
6	521	40.5	1964	9	US-09-864-761-16736		Sequence 16736, A
7	390.4	30.3	3604	10	US-09-764-891-9420		Sequence 9420, Ap
8	390.4	30.3	3604	17	US-10-091-414-288		Sequence 288, App
9	390.4	30.3	3907	9	US-09-823-245A-409		Sequence 409, App
10	272	21.1	3692	18	US-10-723-860-2953		Sequence 2953, Ap
11	272	21.1	5065	18	US-10-723-860-5900		Sequence 5900, Ap

12	272	21.1	5065	18	US-10-723-860-5999	Sequence 5999, Ap
13	272	21.1	5065	18	US-10-723-860-7288	Sequence 7288, Ap
14	256.4	19.9	1144	9	US-09-739-707-1	Sequence 1, Appli
15	214.2	16.6	4839	17	US-10-062-674-1795	Sequence 1795, Ap
16	157	12.2	346	10	US-09-960-706-273	Sequence 273, App
17	157	12.2	346	10	US-09-873-319-165	Sequence 165, App
18	143.4	11.1	2044	9	US-09-764-864-137	Sequence 137, App
19	143	11.1	852	9	US-09-764-864-576	Sequence 576, App
20	142.8	11.1	443	9	US-09-796-692-5430	Sequence 5430, Ap
21	142.8	11.1	443	14	US-10-040-862-5430	Sequence 5430, Ap
22	142.8	11.1	443	17	US-10-057-475B-5430	Sequence 5430, Ap
23	142.8	11.1	443	17	US-10-154-884B-5430	Sequence 5430, Ap
24	142.8	11.1	443	18	US-10-764-324-5430	Sequence 5430, Ap
25	133.8	10.4	350	17	US-10-062-674-56	Sequence 56, Appli
26	113	8.8	282	9	US-09-739-707-3	Sequence 3, Appli
27	105	8.2	3370	17	US-10-104-047-338	Sequence 338, App
28	97.2	7.5	450	10	US-09-918-995-164	Sequence 164, App
29	65.2	5.1	537	10	US-09-918-995-19673	Sequence 19673, A
30	51.6	4.0	640	14	US-10-198-846-13601	Sequence 13601, A
31	51.6	4.0	904	14	US-10-198-846-7085	Sequence 7085, Ap
32	42.6	3.3	497	9	US-09-783-590-5119	Sequence 5119, Ap
33	41.6	3.2	258	17	US-10-641-643-773	Sequence 773, App
34	40.2	3.1	653	14	US-10-184-644-402	Sequence 402, App
35	40.2	3.1	653	14	US-10-184-634-402	Sequence 402, App
36	40	3.1	515	16	US-10-029-386-6506	Sequence 6506, Ap
37	39.4	3.1	1796	18	US-10-425-115-72371	Sequence 72371, A
38	39.2	3.0	289730	18	US-10-719-993-6780	Sequence 6780, Ap
39	38.6	3.0	703	13	US-10-027-632-174146	Sequence 174146,
40	38.6	3.0	703	17	US-10-027-632-174146	Sequence 174146,
41	38.2	3.0	2886	15	US-10-156-761-5473	Sequence 5473, Ap
42	38.2	3.0	9025608	15	US-10-156-761-1	Sequence 1, Appli
43	38	3.0	594	14	US-10-123-155-10	Sequence 10, Appli
44	38	3.0	594	15	US-10-146-731-10	Sequence 10, Appli
45	38	3.0	594	15	US-10-140-472-10	Sequence 10, Appli

ALIGNMENTS

RESULT 1

```
US-08-731-499-5
; Sequence 5, Application US/08731499
; Publication No. US20030148270A1
; GENERAL INFORMATION:
; APPLICANT: GRAY, Joe W.
; APPLICANT: COLLINS, Colin
; APPLICANT: HWANG, Soo-In
; APPLICANT: GODFREY, Tony
; APPLICANT: KOBEL, David
; APPLICANT: KOMMENS, Johanna
; TITLE OF INVENTION: GENES FROM THE 20q13 AMPLICON AND THEIR
; TITLE OF INVENTION: USES
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA: US/08/731,499
; FILING DATE: 16-OCT-1996
; CLASSIFICATION: 435
; PRIORITY APPLICATION NUMBER: US 08/680,395
; FILING DATE: 15-JUL-1996
; ATTORNEY/AGENT INFORMATION:
```

```
/ NAME: Hunter, Tom
/ REGISTRATION NUMBER: 38,498
/ REFERENCE/DOCKET NUMBER: 23070-068910
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (415) 576-0200
/ TELEFAX: (415) 576-0300
/ INFORMATION FOR SEQ ID NO: 5:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 1288 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: cDNA
/ FEATURE:
/ NAME/KEY: -
/ LOCATION: 1..1288
/ OTHER INFORMATION: /note= "cDNA clone 41.1 with homology
/ OTHER INFORMATION: to homeobox T shirt gene from
/ OTHER INFORMATION: Drosophila"
US-08-731-499-5

Query Match      100.0%; Score 1288; DB 8; Length 1288;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1288; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 GAGGGCAGCGAGAAAGGAGAAACCCAGCCCTGGAGCCACATCTGCTCTGAGCAATGGG 60
DB      1 GAGGGCAGCGAGAAAGGAGAAACCCAGCCCTGGAGCCACATCTGCTCTGAGCAATGGG 60

QY      61 TGGCCCTCGCCAAACCAACGCGCCCGGCTGCCATGATCAACCCATCAGCGCCCTGGAG 120
DB      61 TGGCCCTCGCCAAACCAACGCGCCCGGCTGCCATGATCAACCCATCAGCGCCCTGGAG 120

QY      121 TCCGTCCTGAACATCACTTGGGCAAGCCAGCGAGCCCTGGGCTCACCTTCTGCTCC 180
DB      121 TCCGTCCTGAACATCACTTGGGCAAGCCAGCGAGCCCTGGGCTCACCTTCTGCTCC 180

QY      181 AGCCCAAGTTCAAGCACAATTTCCATGTTCCCAAGTCGAATCTCAATGTCAATGGACAAG 240
DB      181 AGCCCAAGTTCAAGCACAATTTCCATGTTCCCAAGTCGAATCTCAATGTCAATGGACAAG 240

QY      241 CCGGTCTTGAGTCTGCTCCCAAGGTCAGCCAGCGTGTCCAGGCGCTACCTGTTTGGAG 300
DB      241 CCGGTCTTGAGTCTGCTCCCAAGGTCAGCCAGCGTGTCCAGGCGCTACCTGTTTGGAG 300

QY      301 AACAGGGATCAGCCCATTCGACCTGACCAAGTCCAAAGGCAAGAACGCCAGTCTCGCAA 360
DB      301 AACAGGGATCAGCCCATTCGACCTGACCAAGTCCAAAGGCAAGAACGCCAGTCTCGCAA 360

QY      361 GCACAATCTTGATGTCTCCACCTCAGAAGCAGCTCTGTCTGACATCGCCGACATGGTC 420
DB      361 GCACAATCTTGATGTCTCCACCTCAGAAGCAGCTCTGTCTGACATCGCCGACATGGTC 420

QY      421 AAAGTCTCTCCCAAGCAACCAACCCAAAGCCAGCTCTCTCTCCAGGGTCCCCCCCATTG 480
DB      421 AAAGTCTCTCCCAAGCAACCAACCCAAAGCCAGCTCTCTCTCCAGGGTCCCCCCCATTG 480

QY      481 AAGCTGGAAATGATGTGAGGCGCTTGGAGATGTCTCCAGTGAAGTCTCAACTTTGCAT 540
DB      481 AAGCTGGAAATGATGTGAGGCGCTTGGAGATGTCTCCAGTGAAGTCTCAACTTTGCAT 540

QY      541 AAAAGAAAAGCGCGGAGTCCAACTGGAATCTCTCAGCATCTTCTGATTTCTACAGGCCAG 600
DB      541 AAAAGAAAAGCGCGGAGTCCAACTGGAATCTCTCAGCATCTTCTGATTTCTACAGGCCAG 600

QY      601 TTGTCCTCGAGCCTCTTCCAGACATCAGAGGCAAAATACCTGTGTCTGATCTGGGCCCA 660
DB      601 TTGTCCTCGAGCCTCTTCCAGACATCAGAGGCAAAATACCTGTGTCTGATCTGGGCCCA 660

QY      661 CAAGAGCGTATGCAATCTTAAGTTTACGGGACTCTCAATGACCACTACAGTCACTGG 720
DB      661 CAAGAGCGTATGCAATCTTAAGTTTACGGGACTCTCAATGACCACTACAGTCACTGG 720
```

```
721 CTGGCCAAACGTCAAGTACAGCTTAGGAAAACGGGGGACAAAATTTCTGAAAAACATG 780
DB      721 CTGGCCAAACGTCAAGTACAGCTTAGGAAAACGGGGGACAAAATTTCTGAAAAACATG 780

QY      781 GACAAAGGCCACCCCATCTTTTATTCAGTGAATCTCCCTCCAGTTTCAAGACCCCTTCT 840
DB      781 GACAAAGGCCACCCCATCTTTTATTCAGTGAATCTCCCTCCAGTTTCAAGACCCCTTCT 840

QY      841 ACCTACATCAGTCACTTAGAATCTCACCTGGGTTCCTCAATGAAGGACATGACCCGCTTG 900
DB      841 ACCTACATCAGTCACTTAGAATCTCACCTGGGTTCCTCAATGAAGGACATGACCCGCTTG 900

QY      901 TCAGTGGACCAAGCAAGGTGGAGCAAGAGATCTCCCGGTATCGTGGGCTCAGAGG 960
DB      901 TCAGTGGACCAAGCAAGGTGGAGCAAGAGATCTCCCGGTATCGTGGGCTCAGAGG 960

QY      961 TCTCAGAAAACATAGTCTCCGAGAGGACACAGACTCTAAATTCAGTGTAAAGTTGTGC 1020
DB      961 TCTCAGAAAACATAGTCTCCGAGAGGACACAGACTCTAAATTCAGTGTAAAGTTGTGC 1020

QY      1021 TGTGCGACATTTGTGAGCAAAACATCGGTAAACCTCCACTAAGCAAAAACGACACCAAG 1080
DB      1021 TGTGCGACATTTGTGAGCAAAACATCGGTAAACCTCCACTAAGCAAAAACGACACCAAG 1080

QY      1081 TCACCCGAAACACCATTTACAGTTTGTAAACAGACGTGGATGAAGAAATAGCTCTGCAGGACG 1140
DB      1081 TCACCCGAAACACCATTTACAGTTTGTAAACAGACGTGGATGAAGAAATAGCTCTGCAGGACG 1140

QY      1141 AATGCTTAGTTTCACTTTCCAGCTTGTAGTCCCTCAGCTGAAACCTTCTTCGTTGCA 1200
DB      1141 AATGCTTAGTTTCACTTTCCAGCTTGTAGTCCCTCAGCTGAAACCTTCTTCGTTGCA 1200

QY      1201 CCATCTGCTTCTGACATTTCACTTGAATCTCTCTGACACACCTGCTCTGAGAAGA 1260
DB      1201 CCATCTGCTTCTGACATTTCACTTGAATCTCTCTGACACACCTGCTCTGAGAAGA 1260

QY      1261 CTGCCAAAAAATAAAAAAAAAAAAAATTC 1288
DB      1261 CTGCCAAAAAATAAAAAAAAAAAAAATTC 1288

RESULT 2
US-10-104-047-373
; Sequence 373, Application US/10104047
; Publication No. US20030236392A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. US20030236392A1e1 full length cDNA
; FILE REFERENCE: H1-A0105
; CURRENT APPLICATION NUMBER: US/10/104,047
; CURRENT FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER:
; PRIOR FILING DATE:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 373
; LENGTH: 3322
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-104-047-373

Query Match      98.2%; Score 1265; DB 17; Length 3322;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1265; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 GAGGGCAGCGAGAAAGGAGAAACCCAGCCCTGGAGCCACATCTGCTCTGAGCAATGGG 60
DB      2058 GAGGGCAGCGAGAAAGGAGAAACCCAGCCCTGGAGCCACATCTGCTCTGAGCAATGGG 2117

QY      61 TGGCCCTCGCCAAACCAACGCGCCCGGCTGCCATGATCAACCCACTCAGCGCCCTGCAG 120
DB      2118 TGGCCCTCGCCAAACCAACGCGCCCGGCTGCCATGATCAACCCACTCAGCGCCCTGCAG 2177
```

QY 121 TCCGTCCTGAACATCACTTGGGCAAGCCAGGAGCCCTTGGCTCACCTTCTGCTCC 180
Db 2178 TCCGTCCTGAACATCACTTGGGCAAGCCAGGAGCCCTTGGCTCACCTTCTGCTCC 2237
QY 181 AGCCCAAGTTCACGACAAATTTTCATGTTCCACAAGTCGAATCTCAATGTCTATGACACAAG 240
Db 2238 AGCCCAAGTTCACGACAAATTTTCATGTTCCACAAGTCGAATCTCAATGTCTATGACACAAG 2297
QY 241 CCGGTCTTGAAGTCTGCTCCACAAAGTTCAGCCAGCGGTTCAGGCGCTACCTGTTTGAAG 300
Db 2298 CCGGTCTTGAAGTCTGCTCCACAAAGTTCAGCCAGCGGTTCAGGCGCTACCTGTTTGAAG 2357
QY 301 AACAGCGATCAGCCCATTTGACCTGACCAAGTTCAAAAGCAGAAAGCCGAGTCTTCGCA 360
Db 2358 AACAGCGATCAGCCCATTTGACCTGACCAAGTTCAAAAGCAGAAAGCCGAGTCTTCGCA 2417
QY 361 GCACAATCTTGTATGTCCTCCACCTCAGAAAGCAGCTCTGTCTGACATCGCGGACATGTC 420
Db 2418 GCACAATCTTGTATGTCCTCCACCTCAGAAAGCAGCTCTGTCTGACATCGCGGACATGTC 2477
QY 421 AAAGTCTCTCCCAAGCCACCCCAAGGCGAGCTCTCTCTCCAGGCTCCCCCCCATG 480
Db 2478 AAAGTCTCTCCCAAGCCACCCCAAGGCGAGCTCTCTCTCCAGGCTCCCCCCCATG 2537
QY 481 AAGCTGGAATGGATGTGAGGCGCTTTGAGGATGTCTCCAGTGAAGTCTCAACTTTGCAT 540
Db 2538 AAGCTGGAATGGATGTGAGGCGCTTTGAGGATGTCTCCAGTGAAGTCTCAACTTTGCAT 2597
QY 541 AAAAGAAAGCCGCGAGTCACTCAACTGGAATCTCAGCATCTCTGATTTCTACAAGCCGAG 600
Db 2598 AAAAGAAAGCCGCGAGTCACTCAACTGGAATCTCAGCATCTCTGATTTCTACAAGCCGAG 2657
QY 601 TTTGCTCTGAGCCCTCTTCCAGCATCAGAGGCGAAATACCTGCTGTCTGATCTGGGCCA 660
Db 2658 TTTGCTCTGAGCCCTCTTCCAGCATCAGAGGCGAAATACCTGCTGTCTGATCTGGGCCA 2717
QY 661 CAAGAGCGTATGAAATCTCTAAGTTTACGGGACTCTCAATGACCACTATCAGTCACTGG 720
Db 2718 CAAGAGCGTATGAAATCTCTAAGTTTACGGGACTCTCAATGACCACTATCAGTCACTGG 2777
QY 721 CTGGCCAAAGTCAAGTACAGCTTAGGAAACAGGCGGGGACAAATTTCTGAAAGCAATG 780
Db 2778 CTGGCCAAAGTCAAGTACAGCTTAGGAAACAGGCGGGGACAAATTTCTGAAAGCAATG 2837
QY 781 GACAAAGCCACCCCATCTTTTATGAGTCACTGTGCTCCAGTTCAGAACCCCTTCT 840
Db 2838 GACAAAGCCACCCCATCTTTTATGAGTCACTGTGCTCCAGTTCAGAACCCCTTCT 2897
QY 841 ACCTACATCAGTCACTTGAATCTCACTGGGTTTCCAAATGAAGGACATGACCCGCTTG 900
Db 2898 ACCTACATCAGTCACTTGAATCTCACTGGGTTTCCAAATGAAGGACATGACCCGCTTG 2957
QY 901 TCAGTGGACCAAGCAAGGTGGAGCAAGATCTCCGGGTATCGTGGCTCAGAGG 960
Db 2958 TCAGTGGACCAAGCAAGGTGGAGCAAGATCTCCGGGTATCGTGGCTCAGAGG 3017
QY 961 TCTCCAGAAAATAAGTTCGCGAAGAGGACACAGACTCTAAATTCAGGTGAAGTTGTC 1020
Db 3018 TCTCCAGAAAATAAGTTCGCGAAGAGGACACAGACTCTAAATTCAGGTGAAGTTGTC 3077
QY 1021 TGTGGCAATTTGTGAGCAAAATCGCGGTAAATCTCACTTAAGCAAAACGACAGCAAG 1080
Db 3078 TGTGGCAATTTGTGAGCAAAATCGCGGTAAATCTCACTTAAGCAAAACGACAGCAAG 3137
QY 1081 TCACCCGAAACCAATTCAGTTTGAACAGCGTGAATGAATAGCTCTGAGGAGC 1140
Db 3138 TCACCCGAAACCAATTCAGTTTGAACAGCGTGAATGAATAGCTCTGAGGAGC 3197
QY 1141 AATGCTTTAGTTTCACTTTCCAGCGCTGGATCCCTCAGCTGAACCCCTTCTTGTGCA 1200
Db 3198 AATGCTTTAGTTTCACTTTCCAGCGCTGGATCCCTCAGCTGAACCCCTTCTTGTGCA 3257
QY 1201 CCATCCTGCTTCTGACATTTGAATCTCACTTCTCTGACACCCCTGGCTCTGAGAAGA 1260

Db 3258 CCATCCTGCTTCTGACATTTGAATCTCACTTCTCTGACACCCCTGGCTCTGAGAAGA 3317
QY 1261 CTGCC 1265
Db 3318 CTGCC 3322

RESULT 3
US-09-764-864-628
; Sequence 628, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT223
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 628
; LENGTH: 2449
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (13)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (2437)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-764-864-628

Query Match 60.0%; Score 773; DB 9; Length 2449;
Best Local Similarity 100.0%; Pred. No. 4.3e-231; Indels 0; Gaps 0;
Matches 773; Conservative 0; Mismatches 0

QY 514 GTCTCCAGTGAAGTCTCAACTTTGTCATATAAAGAAAGCGCGGAGTCCAACTGGAATCCT 573
Db 95 GTCTCCAGTGAAGTCTCAACTTTGTCATATAAAGAAAGCGCGGAGTCCAACTGGAATCCT 154
QY 574 CAGCATCTTCTGATCTTACAAGCCAGTTTGCCTCGAGCCTTCCAGACATCAGAGGCG 633
Db 155 CAGCATCTTCTGATCTTACAAGCCAGTTTGCCTCGAGCCTTCCAGACATCAGAGGCG 214
QY 634 AAATACCTGCTGTCTGATCTGGGCCCCACAGAGGATGCAAAATCTCTAAGTTTACGGGA 693
Db 215 AAATACCTGCTGTCTGATCTGGGCCCCACAGAGGATGCAAAATCTCTAAGTTTACGGGA 274
QY 694 CTCTCAATGACCACTATCAGTCACTGGCTGGCCAAACGTCAGTACCAGCTTAGGAAAACG 753
Db 275 CTCTCAATGACCACTATCAGTCACTGGCTGGCCAAACGTCAGTACCAGCTTAGGAAAACG 334
QY 754 GGCGGGAATAATTTCTGAAAAACATGGACAAAGGCCACCCCATCTTTTATTGCAAGTAC 813
Db 335 GGCGGGAATAATTTCTGAAAAACATGGACAAAGGCCACCCCATCTTTTATTGCAAGTAC 394
QY 814 TGTGCTCCCAAGTTCAGAACCCCTTCTACCTACATCAGTCACTTAGAATCTCACCTGGGT 873
Db 395 TGTGCTCCCAAGTTCAGAACCCCTTCTACCTACATCAGTCACTTAGAATCTCACCTGGGT 454
QY 874 TTCCAAATGAAGGACATGACCCGCTTGTGAGTGGACCAAGCAAGGTTGAGCAAGAG 933
Db 455 TTCCAAATGAAGGACATGACCCGCTTGTGAGTGGACCAAGCAAGGTTGAGCAAGAG 514
QY 934 ATCTCCCGGGTATCGTGGCTCAGAGGTCTCCAGAAACAATAGCTGCCGAAGAGACACA 993
Db 515 ATCTCCCGGGTATCGTGGCTCAGAGGTCTCCAGAAACAATAGCTGCCGAAGAGACACA 574
QY 994 GACTCTAAATTCAGTGTAAAGTTGCTGTGGGACATTTGTGAGCAAAACATCGCGTAAAA 1053
Db 575 GACTCTAAATTCAGTGTAAAGTTGCTGTGGGACATTTGTGAGCAAAACATCGCGTAAAA 634

QY 1054 CTCACCTAAGCAAAACGACAGCAAGTCAACCGAACACCAATTCACAGTTTGTAAACAGAC 1113
DB 635 CTCACCTAAGCAAAACGACAGCAAGTCAACCGAACACCAATTCACAGTTTGTAAACAGAC 694
QY 1114 GTGGATGAAGAATAGCTCTGAGGAGCAATGCTTAGTTTCCACTTTCAGCTCGATCC 1173
DB 695 GTGGATGAAGAATAGCTCTGAGGAGCAATGCTTAGTTTCCACTTTCAGCTCGATCC 754
QY 1174 CTTACACTGAACCTTCTTCTGTTGACCACTCTGCTTCTGACATTTGAATCAATGAAC 1233
DB 755 CTTACACTGAACCTTCTTCTGTTGACCACTCTGCTTCTGACATTTGAATCAATGAAC 814
QY 1234 CTTCTGTACACCTGGCTCTGAGAGACTGCGCAAAACCAAAACCAAAACCAAAAC 1286
DB 815 CTTCTGTACACCTGGCTCTGAGAGACTGCGCAAAACCAAAACCAAAACCAAAAC 867

RESULT 4

US-09-764-864-193
; Sequence 193, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT223
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 192
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 193
; LENGTH: 799
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-864-193

Query Match 59.8%; Score 770.4; DB 9; Length 799;
Best Local Similarity 99.9%; Pred. No. 1.7e-230;
Matches 771; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 514 GTCTCAGTGAAGTCTCAACTTTTGCATATAAAGAAAGCGCGAGTCCAACTGGAAATCT 573
DB 12 GTCTCAGTGAAGTCTCAACTTTTGCATATAAAGAAAGCGCGAGTCCAACTGGAAATCT 71
QY 574 CAGCATCTCTCATCTACAAGCCAGTTTGCCTCGAGCCTCTCCAGACATCAGAGGCG 633
DB 72 CAGCATCTCTCATCTACAAGCCAGTTTGCCTCGAGCCTCTCCAGACATCAGAGGCG 131
QY 634 AAATACCTCTCTGATCTGGGCCCCACAAAGCGGTATGCAATCTCTAAATTTTACGGGA 693
DB 132 AAATACCTCTCTGATCTGGGCCCCACAAAGCGGTATGCAATCTCTAAATTTTACGGGA 191
QY 694 CTCTCAATGACCATATCATAGTCACTGGCTGGCCAAAGTCAAGTACAGCTTAGGAAAAG 753
DB 192 CTCTCAATGACCATATCATAGTCACTGGCTGGCCAAAGTCAAGTACAGCTTAGGAAAAG 251
QY 754 GCGGGGCAAAATTTCTGAABACATGGACAAAGCCACCCATCTTTTATTCAGTGC 813
DB 252 GCGGGGCAAAATTTCTGAABACATGGACAAAGCCACCCATCTTTTATTCAGTGC 311
QY 814 TGTGCTCCAGTTTCAGAACCCCTTCTACTTACATCAGTCACTTAGAATCTCACCTGGGT 873
DB 312 TGTGCTCCAGTTTCAGAACCCCTTCTACTTACATCAGTCACTTAGAATCTCACCTGGGT 371
QY 874 TTCCAAATGAAGACATGACCCGCTTGTGAGTGGACCAAGCAAGAGTGGAGCAAGAG 933
DB 372 TTCCAAATGAAGACATGACCCGCTTGTGAGTGGACCAAGCAAGAGTGGAGCAAGAG 431
QY 934 ATCTCCCGGTATCTGCGCTCAGAGTCTCCAGAAACAATAGCTGCCAAGAGACACA 993
DB 432 ATCTCCCGGTATCTGCGCTCAGAGTCTCCAGAAACAATAGCTGCCAAGAGACACA 491

QY 994 GACTCTAAATTCAGTGAAGTTGTCTGTGGACATTTGTGAGCAAAACATCGGTAAAA 1053
DB 492 GACTCTAAATTCAGTGAAGTTGTCTGTGGACATTTGTGAGCAAAACATCGGTAAAA 551
QY 1054 CTCACCTAAGCAAAACGACAGCAAGTCAACCGAACACCAATTCACAGTTTGTAAACAGAC 1113
DB 552 CTCACCTAAGCAAAACGACAGCAAGTCAACCGAACACCAATTCACAGTTTGTAAACAGAC 611
QY 1114 GTGGATGAAGAATAGCTCTGAGGAGCAATGCTTAGTTTCCACTTTCAGCTCGATCC 1173
DB 612 GTGGATGAAGAATAGCTCTGAGGAGCAATGCTTAGTTTCCACTTTCAGCTCGATCC 671
QY 1174 CTTACACTGAACCTTCTTCTGTTGACCACTCTGCTTCTGACATTTGAATCAATGAAC 1233
DB 672 CTTACACTGAACCTTCTTCTGTTGACCACTCTGCTTCTGACATTTGAATCAATGAAC 731
QY 1234 CTTCTGTACACCTGGCTCTGAGAGACTGCGCAAAACCAAAACCAAAACCAAAAC 1285
DB 732 CTTCTGTACACCTGGCTCTGAGAGACTGCGCAAAACCAAAACCAAAACCAAAAC 783

RESULT 5

US-09-864-761-33203
; Sequence 33203, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aemica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1


```

; SEQ ID NO 33203
; LENGTH: 2012
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC004505.1
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1
; OTHER INFORMATION: NT HIT: AF207880.1, EVALU0.0.00e+00
; OTHER INFORMATION: EST HUMAN HIT: H12950.1, EVALU0.0.00e+00
; OTHER INFORMATION: SWISSPROT HIT: P22265, EVALU0.0.00e-07
US-09-864-761-33203

Query Match          51.0%; Score 656.4; DB 9; Length 2012;
Best Local Similarity 99.8%; Pred. No. 1.4e-194;
Matches 657; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GAGGGCAGCGAAGAGGAAACCCAGCCCTCGAGGCCACATCTGCTCTGAGCAATGGG 60
Db 1347 GAGGGCAGCGAAGAGGAAACCCAGCCCTCGAGGCCACATCTGCTCTGAGCAATGGG 1406

Qy 61 TGCGCCCTCGCAACCAACGCGCCCGCCCTGCGCATGATCAACCACTCAGCGCCCTCGAG 120
Db 1407 TGCGCCCTCGCAACCAACGCGCCCGCCCTGCGCATGATCAACCACTCAGCGCCCTCGAG 1466

Qy 121 TCGTCTCTGAACATCACTTTGGGCAAGCAGGAGGCCCTTGCGCTCACCTTCTCTGCTCC 180
Db 1467 TCGTCTCTGAACATCACTTTGGGCAAGCAGGAGGCCCTTGCGCTCACCTTCTCTGCTCC 1526

Qy 181 AGCCCAAGTTCAAGCACAATTTTCATGTTTCCACAAGTCGAATCTCAATGTCATGGACAAG 240
Db 1527 AGCCCAAGTTCAAGCACAATTTTCATGTTTCCACAAGTCGAATCTCAATGTCATGGACAAG 1586

Qy 241 CCGGTCCTTGAGTCTGCGCTCCACAAGGTCAGCCAGCGTGTCCAGGCGCTACCTCTTTTGAG 300
Db 1587 CCGGTCCTTGAGTCTGCGCTCCACAAGGTCAGCCAGCGTGTCCAGGCGCTACCTCTTTTGAG 1646

Qy 301 AACAGGATCAGGCCATTTGACTGACCAAGTTCAAAAGCAAGAACGCGAGTCTCTCGCAA 360
Db 1647 AACAGGATCAGGCCATTTGACTGACCAAGTTCAAAAGCAAGAACGCGAGTCTCTCGCAA 1706

Qy 361 GCACAATCTTGATGTCCCACTCAGAGCAGCGCTGTCTGTGACATCGCGACATGGTC 420
Db 1707 GCACAATCTTGATGTCCCACTCAGAGCAGCGCTGTCTGTGACATCGCGACATGGTC 1766

Qy 421 AAAGTCCTCCCCAAAGCACCACCCCAAGCAGCGCTCTCTCCAGGGTCCCCCCCATTG 480
Db 1767 AAAGTCCTCCCCAAAGCACCACCCCAAGCAGCGCTCTCTCCAGGGTCCCCCCCATTG 1826

Qy 481 AAGCTGGAATATGATGTGAGCGGCTTTGAGGATGTCTCCAGTGAAGTCTCAACTTTGCAT 540
Db 1827 AAGCTGGAATATGATGTGAGCGGCTTTGAGGATGTCTCCAGTGAAGTCTCAACTTTGCAT 1886

Qy 541 AAAAGAAAGCCGCGAGTCCAACTGGATCCTCAGCATCTTCTGATTTCTACAGGCCAG 600
Db 1887 AAAAGAAAGCCGCGAGTCCAACTGGATCCTCAGCATCTTCTGATTTCTACAGGCCAG 1946

Qy 601 TTTGCTCGAGCCTCTTCCAGACATCAGAGGGCAAAATACCTGCTGTCTGATCTGGGCC 658
Db 1947 TTTGCTCGAGCCTCTTCCAGACATCAGAGGGCAAAATACCTGCTGTCTGATCTGGGCC 2004

RESULT 6
US-09-864-761-16736
; Sequence 16736, Application US/09864671
; Patent No. US2002004876A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USE
; FILE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aemica-X-1

```

;	CURRENT APPLICATION NUMBER: US/09/864,761	
;	CURRENT FILING DATE: 2001-05-23	
;	PRIOR APPLICATION NUMBER: US 60/180,312	
;	PRIOR FILING DATE: 2000-02-04	
;	PRIOR APPLICATION NUMBER: US 60/207,456	
;	PRIOR FILING DATE: 2000-05-26	
;	PRIOR APPLICATION NUMBER: US 09/632,366	
;	PRIOR FILING DATE: 2000-08-03	
;	PRIOR APPLICATION NUMBER: GB 24263.6	
;	PRIOR FILING DATE: 2000-10-04	
;	PRIOR APPLICATION NUMBER: US 60/236,359	
;	PRIOR FILING DATE: 2000-09-27	
;	PRIOR APPLICATION NUMBER: PCT/US01/00666	
;	PRIOR FILING DATE: 2001-01-30	
;	PRIOR APPLICATION NUMBER: PCT/US01/00667	
;	PRIOR FILING DATE: 2001-01-30	
;	PRIOR APPLICATION NUMBER: PCT/US01/00664	
;	PRIOR FILING DATE: 2001-01-30	
;	PRIOR APPLICATION NUMBER: PCT/US01/00669	
;	PRIOR FILING DATE: 2001-01-30	
;	PRIOR APPLICATION NUMBER: PCT/US01/00665	
;	PRIOR FILING DATE: 2001-01-30	
;	PRIOR APPLICATION NUMBER: PCT/US01/00668	
;	PRIOR FILING DATE: 2001-01-30	
;	PRIOR APPLICATION NUMBER: PCT/US01/00663	
;	PRIOR FILING DATE: 2001-01-30	
;	PRIOR APPLICATION NUMBER: PCT/US01/00662	
;	PRIOR FILING DATE: 2001-01-30	
;	PRIOR APPLICATION NUMBER: PCT/US01/00661	
;	PRIOR FILING DATE: 2001-01-30	
;	PRIOR APPLICATION NUMBER: PCT/US01/00670	
;	PRIOR FILING DATE: 2001-01-30	
;	PRIOR APPLICATION NUMBER: US 60/234,687	
;	PRIOR FILING DATE: 2000-09-21	
;	PRIOR APPLICATION NUMBER: US 09/608,408	
;	PRIOR FILING DATE: 2000-06-30	
;	PRIOR APPLICATION NUMBER: US 09/774,203	
;	PRIOR FILING DATE: 2001-01-29	
;	NUMBER OF SEQ ID NOS: 49117	
;	SOFTWARE: Annomax Sequence Listing Engine vers. 1.1	
;	SEQ ID NO 16736	
;	LENGTH: 1964	
;	TYPE: DNA	
;	ORGANISM: Homo sapiens	
;	FEATURE:	
;	OTHER INFORMATION: MAP TO AC004505.1	
;	OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1	
US-09-864-761-16736		
Query Match	40.5%;	Score 521; DB 9; Length 1964;
Best Local Similarity	100.0%;	Pred. No. 4e-152;
Matches 521; Conservative	0; Mismatches	0; Indels 0; Gaps 0;
Qy	1	GAGGGCAGCGAGGAAGGAGAAACCCACGCCCCCTGGAGCCCAACATCTGCTCTGAGCAATGGG 60
Db	1444	GAGGGCAGCGAGGAAGGAGAAACCCACGCCCCCTGGAGCCCAACATCTGCTCTGAGCAATGGG 1503
Qy	61	TGCGCCCTCGCCAAACCAACGCGCCCGGCGCTTGCCATGATCAACCCATCAGCGGCCCTGCAG 120
Db	1504	TGCGCCCTCGCCAAACCAACGCGCCCGGCGCTTGCCATGATCAACCCATCAGCGGCCCTGCAG 1563
Qy	121	TCCGTCCTGACAAATCACTTTGGGCAAGCCACGGAGCCCTTGGCGCTCACCTTCTGCTCC 180
Db	1564	TCCGTCCTGACAAATCACTTTGGGCAAGCCACGGAGCCCTTGGCGCTCACCTTCTGCTCC 1623
Qy	181	AGCCCCAAGTTCAGCAACAATTTCCATGTTTCCAAAGTTCGAATCTCAATGTCAATGTATGGACAAG 240
Db	1624	AGCCCCAAGTTCAGCAACAATTTCCATGTTTCCAAAGTTCGAATCTCAATGTATGGACAAG 1683
Qy	241	CCGGTCTTGAGTCTCTGCTCCACAAGGTGACGAGCGGTGTCAGGGCGCTACCTGTTTGAG 300
Db	1684	CCGGTCTTGAGTCTCTGCTCCACAAGGTGACGAGCGGTGTCAGGGCGCTACCTGTTTGAG 1743

QY 301 AACAGGATCAGCCATTGACCTGACCAAGTCCAAAGCAAGAAAGCCGAGTCTCTGCAA 360
 DB 1744 AACAGGATCAGCCATTGACCTGACCAAGTCCAAAGCAAGAAAGCCGAGTCTCTGCAA 1803
 QY 361 GCACATCTTGATGTGCTCCCACTCAGAGCAGGCTCTGTGACATCCGCGACATGTC 420
 DB 1804 GCACATCTTGATGTGCTCCCACTCAGAGCAGGCTCTGTGACATCCGCGACATGTC 1863
 QY 421 AAAGTCTCTCCCAAGCCACCCCAAGCCAGCCCTCTCTCCAGGGTCCGCCCATG 480
 DB 1864 AAAGTCTCTCCCAAGCCACCCCAAGCCAGCCCTCTCTCCAGGGTCCGCCCATG 1923
 QY 481 AAGCTGGAATGATGTCAGGCGCTTTGAGGATGTCCTCCAG 521
 DB 1924 AAGCTGGAATGATGTCAGGCGCTTTGAGGATGTCCTCCAG 1964
 RESULT 7
 US-09-764-891-9420
 ; Sequence 9420, Application US/09764891
 ; Publication No. US20030077808A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Rosen et al.
 ; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
 ; FILE REFERENCE: PC006
 ; CURRENT APPLICATION NUMBER: US/09/764,891
 ; CURRENT FILING DATE: 2001-01-17
 ; Prior application data removed - consult PALM or file wrapper
 ; NUMBER OF SEQ ID NOS: 10231
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 9420
 ; LENGTH: 3604
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: SITE
 ; LOCATION: (3443)
 ; OTHER INFORMATION: n equals a,t,g, or c
 US-09-764-891-9420

Query Match 30.3%; Score 390.4; DB 10; Length 3604;
 Best Local Similarity 62.4%; Pred. No. 4.6e-111;
 Matches 673; Conservative 0; Mismatches 381; Indels 24; Gaps 3;
 QY 52 AGCAATGGGTGGCGCTTGGCCCAACCAAGCCCGGCTTGCATGATCAACCCACTCAGC 111
 DB 1243 AACAACTGGGATCAATGACCACTCACCAGGAGCTTCTTCAATCAACCGCTGAGC 1302
 QY 112 GCCCTGAGTCCGCTCTGAAACATCACTTTGGGCAAGCCACGAGGCCCTTGGCTCACCT 171
 DB 1303 GCTTTGAGTCCATCATGAACCCACCTGGGCAAGGTGTCAGGCCCTG----- 1353
 QY 172 TCGTGTCCAGCCCAAGTTCAGACCAATTTCAATGTTCCAGATGCAAGTCGAATCTCAATGTC 231
 DB 1354 -----AGTCCCTCGTGGACCGCTGGGATGCTGTACAGATCAGCAACAGCATG 1404
 QY 232 ATGGACAAGCGGTCTTGGATGCTGCTCCACAGGTTCAGCCAGCGTGTCCAGGCGCTAC 291
 DB 1405 CTGGACAAGCGGTGTACCCCGCCACCCCTGTGAAGCAGCGATGCCATCGACCGCTAC 1464
 QY 292 CTGTTTGAACAAGCGATCAGCCATTGACCTGACCAAGTCCAAAGCAAGAAAGCCGAG 351
 DB 1465 TATTATGAACAAGCGACCGACCCATTGACTTAACCAAGTCCAGAAACGCGCTGGTG 1524
 QY 352 TCCTCCGAAGCAAACTCTTGATGTGCTCCCACTCAGAGCAGCGCTCTGTGACATCGCC 411
 DB 1525 TCAGCGGTGGCTGATTCGCTGGGATCACCTCTCGGGAGAGCGCACTCATGACATCTCC 1584
 QY 412 GACATGTCAAAGTCTCTCCCAAGCCACACCCCAAGCGCTCTCTCCAGGGTC 471
 DB 1585 GACATGTCAAAACTCTCAGAGCGCGCTGACGCCCAAGTCTCTCAGCGCTCCACAGT 1644
 QY 472 CCCCCATGAAGCTGGAATGGATGTGAGCGCTTTGAGGATGTCCTCCAGTGAAGTCTCA 531

DB 1645 TCAGA---GAAGTCCGATGCTGATGCGACAGCTTTGAGGAGCGTTGGACGAGCTGCA 1701
 QY 532 ACTTTGCAATAAGAAAGGCGGAGTCCAACTGGAATCCTCAGCATCTTCTGATCTTA 591
 DB 1702 CCGGTCCAAGAGGAAGGCGGAGTCCAACTGGAACCCGACGACCTTCTCATCTCG 1761
 QY 592 CAAGCCAGTCTTGGCTCGAGCCTCTTCCAGACATCAGAGGGCAAAATACCTGTGTCTGAT 651
 DB 1762 CAGGCCAGTCTGCTCTGAGCTTGGGGAGACCAAGAGGCAAGTACATCATGTCCGAC 1821
 QY 652 CTGGGCCCAAGAGCGGTATGCAAACTCTAAAGTTTACGGGACTCTCAATGACCACTATC 711
 DB 1822 TTGGGCCCGCAGAGAGGGTGCACTCTCGAAGTTTACTGGGCTCTCCATGACCCCATC 1881
 QY 712 AGTCACTGGCTGGCCAAAGTCAAGTACAGCTTAGGAAAACGGCGGGGCAAAATTTCTG 771
 DB 1882 AGCCACTGGCTGGCCAAAGTGAAGTACCAAGTTGAGGAGGACAGGGGGAAACGAAATTTCTA 1941
 QY 772 AAAAAATGACAAAGGCCACCCCATCTTTTATTCAGTGAAGTCTGCTCCCTCCAGTTTCA 831
 DB 1942 AAGAACCTGGACACAGGGGATCTGTTTCTTTTGGCAAGATTTGCTCTCAGTTTCA 2001
 QY 832 ACCCTTCTACCTACATCAGTCACTTAGAATCTCACCTGGTTTCCAAATGAAGGACATG 891
 DB 2002 ACTGCTTCTACATACATAAGTCACTTTGGAGACACACTTGGGCTTCAAGCTGAAGGATCTC 2061
 QY 892 ACCCGCTTCTAGTGGACCAAGCAAGGTGGAGCAAGATCTCCCGGGTATCGTCG 951
 DB 2062 TCCAAAGCTGCCACT---CAATCAGATTCAAGAACACGAGAATGTTTCGAAAGTCTCTCACC 2118
 QY 952 GCTCAGAGGTCTCCAGAAACAATAGTCTGCCAAGAGGACACAGACTCTTAAATTTCAAGTGT 2178
 DB 2119 AACAAAATCTGGGCGCCACTGGGGGCCACCCAGGAGAGACTTGGGCTCCACATTTCAATGT 2178
 QY 1012 AAGTTGTGCTGCGGACATTTGTCGACAAATGTCGGTAAAACTCCACTAAGCAAAACG 1071
 DB 2179 AAGCTCTGCAACCGGACTTTTGGCAAGCAAGCAGCAGTCAAACTGCACTTAGTAAGACC 2238
 QY 1072 CACAGCAAGTCAACCGAACCACTTCAAGTTTGTAAACAGCTGATGAAGATAGC 1129
 DB 2239 CACGCAAGTCTCCGAGGACCACTGATCTATGACTGAGTTGGAGAAACAGTAGC 2296
 RESULT 8
 US-10-091-414-288
 ; Sequence 288, Application US/10091414
 ; Publication No. US20030224461A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Rosen et al.
 ; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
 ; FILE REFERENCE: P1116C1
 ; CURRENT APPLICATION NUMBER: US/10/091,414
 ; CURRENT FILING DATE: 2002-03-07
 ; Prior Application removed - See File Wrapper or Palm
 ; NUMBER OF SEQ ID NOS: 392
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 288
 ; LENGTH: 3604
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: misc feature
 ; LOCATION: (3443)
 ; OTHER INFORMATION: n equals a,t,g, or c
 US-10-091-414-288
 Query Match 30.3%; Score 390.4; DB 17; Length 3604;
 Best Local Similarity 62.4%; Pred. No. 4.6e-111;
 Matches 673; Conservative 0; Mismatches 381; Indels 24; Gaps 3;
 QY 52 AGCAATGGGTGGCGCTTGGCCCAACCAAGCCCGGCTTGCATGATGAAGTGAAGTCAAGC 111

Db 1243 AACAACTGGGATCATATGACCACTACCGGAGCCTTCTTTCATCAACCGCTGAGC 1302
Qy 112 GCCCTGAGTCCGCTCTGAACAAATCACTTGGGCAAGCCACGAGCCCTTGGCTCACT 171
Db 1303 GCTTTGCAATCCATCATGAACACCCACCTGGGCAAGGTGTCCAAAGCCGCTG----- 1353
Qy 172 TCCTGCTCCAGCCCAAGTTCAAGACAAATTTCCATGTTCCCAAGTCGAATCTCAATGTC 231
Db 1384 -----AGTCCCTCGTGACCGCTGGCGATGCTGTACAGATCAGCAACAGCATG 1404
Qy 232 ATGGACAAGCCGCTTCTGAGTCTGCTCCCAAGGTCAGCCAGCGGTGTCCAGGCGCTAC 291
Db 1405 CTGGACAAGCCGCTGTACCCGCCACCCCTGTGAAGCAGCGCGATGCCATCGCCGCTAC 1464
Qy 292 CTGTTTGAAACAGCGATCAGCCATTTGACCTGACCAAGTCCAAAGCAAGAAAGCCGAG 351
Db 1465 TATTATGAAACAGCGACGACCCATTTGACTTAAACCAAGTCCAAAGCAAGCCGCTGGT 1524
Qy 352 TCCTCGCAAGCACAAATCTTGATGTTCCCACTCAGAGCAGCTCTGCTGACATGCC 411
Db 1525 TCCAGCGTGGCTGATTCGGTGGCATCACTCTGCGGGAGAGCGCACTCATGGACATCTCC 1584
Qy 412 GACATGGTCAAAAGTCTCTCCCAAGGCCACCCCAAGCCAGCCCTCTCTCCAGGGTC 471
Db 1585 GACATGGTGAACAACTCACAGGCGCTGAGCGCCAAAGTCTTCCAGCCCTCCACAGT 1644
Qy 472 CCCCCCATGAAGCTGGAATGATGTGAGGCGCTTTGAGGATGTCTCCAGTGAATCTCA 531
Db 1645 TCAGA---GAACTCCGATGCTGATGGCAGCAGCTTTGAGGAGCGTTGAGCAGCTGCA 1701
Qy 532 ACTTTGCAATAAGAAAGGCGGCGAGTCCCACTGGAATCTTCAGCATCTTCTGATTTCA 591
Db 1702 CCGGTCCCAAGAGAAAGGCGGCGAGTCCCACTGGAACCCCGCAGCCTTCTCATCTG 1761
Qy 592 CAAGCCCAAGTTGCTCGAGCTCTTCCAGACATCAGAGGGCAAAATACCTGCTGCTGAT 651
Db 1762 CAGGCCAGTTCGCTTCAGCTTGGGGAGACCAAGAGGCGAAGTATCATATGTCGAC 1821
Qy 652 CTGGCCCCCAAGAGCGTATGCAATCTCTAAGTTTACGGGATCTCAATGACCACTATC 711
Db 1822 TTGGCCCCGAGGAGGCTGACATCTCGAAGTTTACTGGGCTCTCCATGACCACTATC 1881
Qy 712 AGTCACTGGCTGGCCCAAGTCAAGTACCACTTAGGAAACCGGGGGGACAAATTTCTG 771
Db 1882 AGCCACTGGCTGGCCCAATGTGAAGTACCAGTTGAGGAGGACAGGGGGGAAACGAAATTCCTA 1941
Qy 772 AAAACATGGACAAAGGCCACCCCATCTTTTATTCAGTGAAGTGTCTCCAGTTTCA 831
Db 1942 AAGAACCTGGACACAGGGGATCTCTGTTTCTTTTGCACGATTTGGCTCTCAGTTTCA 2001
Qy 832 ACCCTTCTACCTACATCAGTCACTTAGAATCTCACTGGGTTTCCAAATGAAGGACATG 891
Db 2002 ACTGCTTCTACATACATAAGTCAATTTGGAGACACACTTTGGGCTTCAGCTGAAGATCTC 2061
Qy 892 ACCCGCTTGTCACTGGACCAAGCAAGTGGAGCAAGATCTCCGGGTATCTGTCG 951
Db 2062 TCCAAAGTGGCACT---CAATCAGATTCAAGAACAGCAGAAATGTTTCGAAAGTCTCTACC 2118
Qy 952 GCTCAGAGGTCTCCAGAAACAAATAGCTGCGAAGAGGACACAGACTCTTAAATTCAGGT 1011
Db 2119 AACAAATCTGGGGCCCTAGGGGCGCCACCGAGGAAGACTTGGGCTCCACATTCATATG 2178
Qy 1012 AAGTTGTCTGTGCGACATTTTGTAGCAAAACATGCGGTAAATCTCACCTAAGCAAAACG 1071
Db 2179 AAGCTCTCAACCGGACTTTTTCGAGCAAGCAGCAGTCAAACTGACCTTAGTAAGACC 2238
Qy 1072 CACAGAGTCAACCGAACCACTTCAAGTTTGTAAACAGAGCTGGATGAAGATAGC 1129
Db 2239 CACGCAAGTCTCCCGAGGACCACTGATCTATGTGACTGAGTTGGAGAAACAGTAGC 2296

RESULT 9

US-09-823-245A-409

; Sequence 409, Application US/09823245A
; Publication No. US20020039760A1
; GENERAL INFORMATION:
; APPLICANT: Wong, Gordon G.
; APPLICANT: Clark, Hilary
; APPLICANT: Fechtel, Kim
; APPLICANT: Agostino, Michael J.
; APPLICANT: Howes, Steven H.
; APPLICANT: Resnick, Richard J.
; APPLICANT: Gulukota, Kamalakhar
; APPLICANT: Graham, James R.
; APPLICANT: Genetics Institute, Inc.
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS
; FILE REFERENCE: GIN 6401
; CURRENT APPLICATION NUMBER: US/09/823,245A
; CURRENT FILING DATE: 2001-03-29
; PRIOR APPLICATION NUMBER: 60/194,941
; PRIOR FILING DATE: 2000-04-06
; NUMBER OF SEQ ID NOS: 631
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 409
; LENGTH: 3907
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 3870
; OTHER INFORMATION: n = a,t,g, or c
US-09-823-245A-409

Query Match 30.3%; Score 390.4; DB 9; Length 3907;
Best Local Similarity 62.4%; Pred. No. 4.8e-111;
Matches 673; Conservative 0; Mismatches 381; Indels 24; Gaps 3;
Qy 52 AGCAATGGTGGCGCTCGCCCAACCCAGCCCGCCCTGCCATGTCATCAACCCACTCAGC 111
Db 2425 AACAACTGGGATCATATGACCACTACCGGAGCCTTCTTTCATCAACCGCTGAGC 2484
Qy 112 GCCCTGAGTCCGCTCTGAACAAATCACTTGGGCAAGCCACGAGCCCTTGGCTCACT 171
Db 2485 GCTTTGAGTCCATCATGAACACCCACCTCGGCAAGGTGTCCAAAGCCCGT----- 2535
Qy 172 TCCTGCTCCAGCCCAAGTTCAAGCACAAATTTCCATGTTCCACAAAGTCAATCTCAATGTC 231
Db 2536 -----AGTCCCTCGCTGGACCCCGCTGGCGATGCTGTACAAAGATCAGCAACAGCATG 2586
Qy 232 ATGACCAAGCCGCTCTTGGTCTTCCCTCCACAAAGTCAGCCAGCTGTCCAGGCGCTAC 291
Db 2587 CTGGACAAGCCGCTGTATCCCGCCACCCCTGTGAAGAGCGCGATGCCATCGCCGCTAC 2646
Qy 292 CTGTTTGAGAACAGCGATCAGCCCAATTTGACCTGACCAAGTCCAAAGCAAGAAAGCCGAG 351
Db 2647 TATTATGAAACAGCGACCGACCCCATTTGACTTAAACCAAGTCCAAAGCAAGCCGCTGGTG 2706
Qy 352 TCCTGCGCAAGCACAAATCTTGTATGTCCTCCCTCAGAGCAGCTCTGCTGACATCGCC 411
Db 2707 TCCAGCGTGGCTGATTCGGTGGCATCACCTCTGCGGGAGAGCGCACTCATGGACATCTCC 2766
Qy 412 GACATGGTCAAGTCTCTCCCAAGCCACCCCAAGCCAGCCCTCTCTCCAGGGTTC 471
Db 2767 GACATGGTGAACAACTTCAGAGCGCTGAGCCCAAGTCTTCCAGCCCTCCACAGTT 2826
Qy 472 CCCCCCATGAAGCTGGAATGATGTCAAGGCGCTTTGAGGATGTCTCCAGTGAAGTCTCA 531
Db 2827 TCAGA---GAAAGTCCGATGCTGATGGCAGCAGCTTTGAGGAGCGTTGGAGCGTGTCA 2883
Qy 532 ACTTTCATTAAGAAAGGCGCGAGTCCAACTGGAATCTTCAGCATCTTCTGATTTCTA 591
Db 2884 CCGGTCCCAAGAGAGAGGCGCGAGTCCAACTGGAAACCCGCGAGCACCTTCTCTCACTCTG 2943
Qy 592 CAAGCCCAAGTTTGGCTCGAGCTCTTCCAGACATCAGAGGGGCAAAATACCTGCTGTCTGAT 651
Db 2944 CAGGCCCAAGTTTGGCTCGAGCTTGGGGAGAGCAACAGAGGGCAAGTATCATATGTCGGAC 3003

```

QY 652 CTGGGCCCAAGAGCGTATGAAATCTCTAAGTTTACGGGACTCTCAATGACCACTATC 711
Db 3004 TTGGGCCCGCAGGAGGGGTGCATCTCGAAGTTTACTGGGCTCTCCATGACCACTATC 3063
QY 712 AGTCACTGCTGGCCCAACGCTCAAGTACCAAGCTTTAGGAAAACGGGGGACAAAATTTCTG 771
Db 3064 AGCCACTGGCTGGCCCAATGTGAAGTACCAAGTTGAGGAGACAGGGGGAACGAAATTCCTA 3123
QY 772 AAAACATGGACAAGGCCACCCCATCTTTTATTGAGTACAGTGTGCTCCAGATTGCA 831
Db 3124 AAGAACCCTGGACACAGGGCATCTGTGTTTCTTTTGAACGATTTGTCCTCTCAGTTTCA 3183
QY 832 ACCCTTCTACCTACATCAGTCACTTAGAATCTCACTGGGTTTCCAAATGAAGACATG 891
Db 3184 ACTGCTTCTACATACATAGTCACTTTGGAGACACACTTGGGCTTCAAGCTGAAGATCTC 3243
QY 892 ACCCGCTTGTAGTGGACCAAGCAAGCTGGAGCAAGAGATCTCCCGGCTATCGTGG 951
Db 3244 TCCAAAGCTGCCACT---CAATCAGATTCAAGAACAGCAGATGTTTCGAAAGTCTCTCACC 3300
QY 952 GCTCAGAGTCTCCAGAAACATAGCTGCGGAGAGGACACAGACTCTAAATTCAGTGT 1011
Db 3301 AACAAAACCTCTGGGCCCACTGGGGGCCACCGAGGAAGACTTGGGCTCCACATTTCCAATGT 3360
QY 1012 AAGTTGTGCTGTCGGACATTTGTGAGCAAAACATGCGGTAAACTCCACCTAAGCAAAACG 1071
Db 3361 AGCTCTGCAACCGGACTTTTGGGACAGACGACGAGTCAAACTGACCTTTAGTAGACC 3420
QY 1072 CACAGCAAGTACCCGAAACCACTTACAGTTTGTAAACAGAGCTGGATGAAGAAATAGC 1129
Db 3421 CACGGCAAGTCTCCGAGGACCACCTGATCTATGTGACTGAGTTGAGAAACAGTAGC 3478

RESULT 10
US-10-723-860-2953
; Sequence 2953, Application US/10723860
; Publication No. US20040253606A1
; GENERAL INFORMATION:
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsburg, Wendy M.
; APPLICANT: Zlotnik, Albert
; TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &
; TITLE OF INVENTION: Methods for Screening for Soft Tissue Sarcoma Modulators
; FILE REFERENCE: 05882.0193.NPUS01
; CURRENT APPLICATION NUMBER: US/10/723,860
; PRIOR FILING DATE: 2003-11-26
; PRIOR FILING DATE: 2002-11-26
; NUMBER OF SEQ ID NOS: 8393
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2953
; LENGTH: 3692
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-723-860-2953

Query Match 21.1%; Score 272; DB 18; Length 3692;
Best Local Similarity 62.2%; Pred. No. 6.2e-74;
Matches 469; Conservative 0; Mismatches 270; Indels 15; Gaps 2;

QY 376 TCCCCACCTCAGAACGACGCTGTCTGATCGCGACATGTCGATGATGTCCTCCCAAA 435
Db 2505 TCGCCGCTAGCGGAGATGCTTGTAGATATATCCGATATGCTGAAGACTTGACAGAG 2564
QY 436 GGCACACCCCAAGCCAGCTCTCTCCAGGTCCTCCCGGATGAGCTGGAATGGAT 495
Db 2565 AGCCACACGTCAAAATCTCCACTCTCTCCAGCATCTCC-----GAGAAGTCTGACATT 2618
QY 496 GTGAGGGCTTTGAGGATGCTCCAGTGAAGTCTCACTTTGATATTAAGAAAGGCGCG 555
Db 2619 GACGGGGCCACTCTGAGGAGGCTGAGGAGTCGACCCGCCAGAGAGGAGGCGCGC 2678

```

```

QY 556 CAGTCAACTGGAATCTCTCAGCATCTCTGATTTCTAAGCCCAAGTTTGCCTCGAGCCTC 615
Db 2679 CAGTCAACTGGAATCTCTCAGCATCTCTGATTTCTCAGGCCCAAGTTTGCCTCGAGCCTC 2738
QY 616 TTCCAGACATCAGAGGGCAAAATACCTGCTGTGTGATCTTGGGGCCCAAGAGAGGTATGAA 675
Db 2739 CGGCAGACCTCAGAGGGGAAGTACATCATGTCTGAGCTTGAGCCCGCCAGAGGGATGCAT 2798
QY 676 ATCTCTAAGTTTACCGGACTCTCAATGACCATATCAGTCACTGCTGCTGCCCAACGTCAG 735
Db 2799 ATCTCTAAGTTTACCGGACTCTCAATGACCATATCAGTCACTGCTGCTGCCCAACGTCAG 2858
QY 736 TACCAGCTTAGGAAAACGGGGGAGCAAAATTTCTGAAAAACATGACAAAAGGCCACCC 795
Db 2859 TACCAGCTTCGAAAGCAGGTGAGAACAAAGTTCTCAAAAACCTTGGACACTGGCCACCCC 2918
QY 796 ATCTTTTATTCAGTGTGACTGTGCTCCAGATTTCAGAACCCCTTTCTACCTACATCAGTAC 855
Db 2919 GTCTTCTTTTGAACGATTTGTCGTCCTCCAAATCAGGACTCTCTTCCACGTACATCAGTAC 2978
QY 856 TTAGATCTCACTCGGTTTCCAAATGAAGGACATGACCCGCTTGTCACTGAGCACAGCAA 915
Db 2979 CTAGAGTCACTTTAGGCTTCCGGCTACGGGACTTATCCAAACTGTCCACGAAACAGATT 3038
QY 916 AGCAAGGTGGAGCAAGAGATCTCCGGGTATCGTCCGGCTCAGAGGTCTCCAGAAACAAATA 975
Db 3039 AACA-----GTCAGATAGCACAAACCAAGTCACCGTCAGAAAAAATGGTACGTCC 3089
QY 976 GTGCGGAAAGAGGACACAGACTCTAAATTCAGTGTAAAGTTGTGTGTCGGACATTTGTG 1035
Db 3090 TCCCCCGAGGAAGACTTGGGGACTTCTCTATCAGTGCAAACTTTGCAATCGGACCTTTGCC 3149
QY 1036 ASCAAACATGCGGTAAACTCCACCTAAGCAAAACGACAGCAAGTTCACCCGAAACACCAT 1095
Db 3150 AGCAAGCAGCTGTAAACTTCACCTTAGCAAAACACACGGGAATCTCCGGAAGACCAC 3209
QY 1096 TCACAGTTTGTAAACAGACGTCGATGAAGAAATAGC 1129
Db 3210 CTTCTGTATGTCTCTGAGTTAGAGAGCAGTAGC 3243

RESULT 11
US-10-723-860-5900
; Sequence 5900, Application US/10723860
; Publication No. US20040253606A1
; GENERAL INFORMATION:
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsburg, Wendy M.
; APPLICANT: Zlotnik, Albert
; TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &
; TITLE OF INVENTION: Methods for Screening for Soft Tissue Sarcoma Modulators
; FILE REFERENCE: 05882.0193.NPUS01
; CURRENT APPLICATION NUMBER: US/10/723,860
; PRIOR FILING DATE: 2003-11-26
; PRIOR FILING DATE: 2002-11-26
; NUMBER OF SEQ ID NOS: 8393
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 5900
; LENGTH: 5065
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (520)...(614)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4376)...(4392)
; OTHER INFORMATION: n is a, c, g, or t
US-10-723-860-5900

Query Match 21.1%; Score 272; DB 18; Length 5065;

```

Best Local Similarity 62.2%; Pred. No. 7.1e-74; Matches 469; Conservative 0; Mismatches 270; Indels 15; Gaps 2;	
Qy	376 TCCCCACCTCTAGAACGACGCTCTCTGTGACATCGCCGACATGGTCAAAAGCTCTCCCAAA 435
Db	2644 TCGCGCTACGCGAATGCTTGTTCAGATATATCCGATATATGCTGAAGAATCTTACACAG 2703
Qy	436 GCCAACACCCAAAGCCAGCCTCTCTCTCAGGGTCCCGCCCATGAAGCTGGAATCGAT 495
Db	2704 AGCCACACGTCAAAATCTCCACCTCTTCCAGCATCTCC-----GAGAAGTCTGACATT 2757
Qy	496 GTCAGGCGCTTTGAGGATGTCTCCAGTGAAGTCTCAACTTTTGCATATAAAGAAAGGCCGG 555
Db	2758 GACGGGGCCACTCTGGAGGAGGCTTGAGGAGTCTGACGCCGCCAGAGAGAAGGGCGCC 2817
Qy	556 CAGTCCAACTGGAATCTCTCAGCATCTTCTGTATTCTACAAGCCCAAGTTTGCTCTCAGGCTC 615
Db	2818 CAGTCAAATGGAACCCCGACACTCTGTATCTCTCAGGCCCAAGTTTGGCGCAGGCTC 2877
Qy	616 TTCAGACATCAGAGGGCAATACTCTGTCTGATCTGGGCCCAAGAGCGTATGCAA 675
Db	2878 CGGCAGACCTCAGAAGGGAAGTACATCATGTTCAGACCTGAGCCCCCAGGAGCGGATGCAT 2937
Qy	676 ATCTCTAAGTTTACGGGACTCTCAATGACCACTATCAGTCACTGGCTGGGCCAAGCTCAAG 735
Db	2938 ATCTCCAGGTTTACCGGGCTGTCCATGACCACCATCAGCCACTGGCTGGCCACAGTGAAA 2997
Qy	736 TACCAGCTTTAGGAAAACGGGCGGACAAAATTTCTGAAAACATGGAACAAAGGCCACCCC 795
Db	2998 TACCAGCTTCGAAGGACAGGTGGAAACAAAGTTCTCTAAAACCTTGGACACTGGGCACCCC 3057
Qy	796 ATCTTTTATTGCAGTGACTGTGCCCTCCAGTTTCAGAACCCCTTCTACTACATCAGTCAAC 855
Db	3058 GTCTCTTTTGTACGATTGTGCGTCCCAATCAGGACTCTTTCACGTATCATCAGTCAAC 3117
Qy	856 TTAGAATCTCACTGGGTTTCCAAATGAAGACATGACCGCGTTGTTCAGTGGACCAAGCAA 915
Db	3118 CTAGAGTCACACTTAGGCTTCGGGCTACGGGACTTATCCAAATGTCTCCACCGAACAGATT 3177
Qy	916 AGCAAGGTGGACAAAGATCTCCGGGTATCGTCCGCTCAGAGTCTCAGAAACAATA 975
Db	3178 AACA-----GTCAGTAGCAACAACCAAGTCACCGTTCAGAAAAATGGTGAAGCTCC 3228
Qy	976 GCTCCGAAGGACACAGACTCTAAATTTCAAGTGTAAAGTTGTCTGTCCGACATTTGTG 1035
Db	3229 TCCCCGAGGAAACCTTGGGACTTCCCTATCAGTGCAAAACCTTTGCAATTCGAGCTTTGCC 3288
Qy	1036 AGCAAAACATGCGGTAAACCTCCACTAAGCAAAAACGACACGAAGTCAACCGCAACCAT 1095
Db	3289 AGCAAGCACGCTGTAAACTTCACCTTAGCAAAAACACACGGGAATCTCCGGAAGACCAC 3348
Qy	1096 TCACAGTTTGTAAACAGAGCTGGATGAAGAATPAGC 1129
Db	3349 CTCTCTATGTCTCTGAGTTTAGAAGCAGTAGC 3382

RESULT 12
US-10-723-860-6999
; Sequence 6999, Application US/10723860
; Publication No. US20040253606A1
; GENERAL INFORMATION:
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsburg, Wendy M.
; APPLICANT: Zlotnick, Albert
; TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &
; TITLE OF INVENTION: Methods for Screening for Soft Tissue Sarcoma Modulators
; FILE REFERENCE: 05882.0193.NPUS01
; CURRENT APPLICATION NUMBER: US/10/723,860
; CURRENT FILING DATE: 2003-11-26
; PRIOR APPLICATION NUMBER: 60/429,739
; PRIOR FILING DATE: 2002-11-26
; NUMBER OF SEQ ID NOS: 8393
; SOFTWARE: PatentIn version 3.2

```

; SEQ ID NO 6999
; LENGTH: 5065
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (520)..(614)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4376)..(4392)
; OTHER INFORMATION: n is a, c, g, or t
US-10-723-860-6999

Query Match      21.1%; Score 272; DB 18; Length 5065;
Best Local Similarity 62.2%; Pred. No. 7.1e-74;
Matches 469; Conservative 0; Mismatches 270; Indels 15; Gaps 2;

Qy 376 TCCCCACCTCAGAAGCAGCGTCTGTCTGACATCGCCGACATGGTCAAAGTCTCTCCCCAAA 435
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Dy 2644 TCGCCGCTACGCGAGAATGCTTGTTCAGATATATCCGATATGCTGAAGAATCTTGACAGAG 2703
Qy 436 GCACACACCCCAAGCCAGCGCTCTCTCTCCAGGGTCCGCCCATGAAGCTGGAAATGGAT 495
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Dy 2704 AGCCACAGTCAAAATCCTCACTCTCTCCAGCATCTCC-----GAGAAGTCTGACATT 2757
Qy 496 GTCAGGCGCTTTGAGGATGTCTCCAGTGAAGTCTCAACTTTTGCATAAAGAAAGGCGCG 555
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Dy 2758 GACGGGCGCACTCTGGAGGAGGCTGAGGAGTTCGACGCCGCCGCCAGAGAGGAGGCGCG 2817
Qy 556 CAGTCCAACTGGAAATCTTCAGATCTTCTGATTTCTACAGCCAGTTTGGCTTCGAGCCTC 615
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Dy 2818 CAGTCAAACTGGAAACCCCGAGCACCTCTGATCTCCAGGCCAGTTTGGCGGCGAGCCTC 2877
Qy 616 TTCCAGACATCAGAGGGCAATACCTGCTCTGTGATCTGGGGCCACAGAGCGTATGCAA 675
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Dy 2878 CGGCGAGACCTCAGAAGGGAATGATCATGTGACACTGTAGCCCCCAGGAGCGGATGCAT 2937
Qy 676 ATCTCTAAGTTTACGGGACTCTCAATGACCACATATCAGTCACTGGCTGGCCAAAGCTCAAG 735
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Dy 2938 ATCTCAGGTTACCGGGCTGTCCATGACCACCACCATCAGCCACTGGCTGGCCAAAGCTGAAA 2997
Qy 736 TACCAGTTTAGGAAACGCGGGCGGGAACAAATTTCTGAAAAACATGGACAAAGGCCACCCC 795
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Dy 2998 TACCAGCTTCGAAGGACAGGTGGAAACAAAGTTCTCTCAAAAACCTTGGACACTGGCCACCCC 3057
Qy 796 ATCTTTTATTTGACGTGACTGTGCTCCAGTTCAGAACCCCTTCTACCTACATCAGTCAAC 855
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Dy 3058 GTCTCTCTTTTGTAAACGATTTGGGTGCCAAATCAGGACTCTCTTCCAGTACATCAGTCAAC 3117
Qy 856 TTAGAATCTCACCTGGGTTTTCAAATGAAGGACATGACCCGCTTGTTCAGTGGGACACAGCAA 915
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Dy 3118 CTAGAGTCACTTAGCTTTCCGGCTACGGGACTTATTCAAAATGTCTCCACCGAACAGATT 3177
Qy 916 AGCAAGGTGGACGAAGAGATCTCCCGGGTATCGTCCGCTCAGAGGTTCTCAGAAAAACAATA 975
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Dy 3178 AACA-----GTCAGATAGCACAAACCAAGTCCACCGTCAGAAAAAATGGTGAAGCTCC 3228
Qy 976 GCTGCCGGAAGAGACACAGACTCTAAATTCAGTGTGAAGTTGTGCTGTCGACATTTGTG 1035
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Dy 3229 TCCCGCGAGGAAGACTGGGGGACTTCTATTCAGTGTCAAACTTTTGCAATCGGACCTTTGCC 3288
Qy 1036 AGCAACATCGGTTAAACTCCACCTAAGCAAAACGCACAGCAAGTCCACCCGCAACACCAT 1095
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Dy 3289 AGCAAGCACCGCTGTAAACTTTCACTTTAGCAAAACACACGCGGAATCTCCGGAAGACAC 3348
Qy 1096 TCACAGTTTGTAAACAGACGTGGATGAAGAATAGC 1129
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Dy 3349 CTTCCTGTATGCTCTGATTTAGAGAGCAGTAGC 3382

RESULT 13
US-10-723-860-7288

```

; Sequence 7288, Application US/10723860
; Publication No. US20040253606A1
; GENERAL INFORMATION:
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsburg, Wendy M.
; APPLICANT: Zlotnik, Albert
; TITLE OF INVENTION: Methods for Diagnosis of Soft Tissue Sarcoma, Compositions &
; TITLE OF INVENTION: Methods for Screening for Soft Tissue Sarcoma Modulators
; FILE REFERENCE: 05882.0193.NPUS01
; CURRENT APPLICATION NUMBER: US/10/723,860
; CURRENT FILING DATE: 2003-11-26
; PRIOR APPLICATION NUMBER: 60/429,739
; PRIOR FILING DATE: 2002-11-26
; NUMBER OF SEQ ID NOS: 8393
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 7288
; LENGTH: 5065
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (520)...(614)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4376)...(4392)
; OTHER INFORMATION: n is a, c, g, or t
US-10-723-860-7288

Query Match 21.1%; Score 272; DB 18; Length 5065;
Best Local Similarity 62.2%; Pred. No. 7.1e-74;
Matches 469; Conservative 0; Mismatches 270; Indels 15; Gaps 2;

QY 376 TCCCGACCTCAGNAGCAGCTCTGTCTGACATCGCGCATGTCCTCCCAAA 435
DB 2644 TCGCCCTCAGCGAGAATGCTTGTGAGATATATCGCATGTCGAGAACTTGACAGAG 2703

QY 436 GCCACCAACCCCAAGCCAGCTCTCTCCAGGGTCCCCCATGAGCTGGAATGGAT 495
DB 2704 AGCCACAGCTCAAAATCTCTCCACTCTTCCAGCATCTCC-----GAGAGTGTGACAT 2757

QY 496 GTCCAGCGCTTTGAGATGTCCTCCAGTGAAGTCTCAACTTTGCATAAAAAGAGCCGG 555
DB 2758 GACGGGGCACTCTGGAGGAGGTGAGGAGTCGACGCCGCCCAGAGAGGAGGCCCG 2817

QY 556 CAGTCAACTGGATCTCTGAGATCTTCTGATTTACAGCCCACTTTGCTCGAGCCTC 615
DB 2818 CAGTCAAACTGGAACCCCAAGCACTCTCTGATCTCCAGGCCCACTTTGCGCCAGCCTC 2877

QY 616 TTCCGACATCAGAGGCAAAATACCTGTCTGTCTGATCTGGGCCCAAGAGCGTATGCAA 675
DB 2878 CGCGACCTCAGAGGAAAGTACATCTGTGAGCTGAGCTGAGCCCCCAGAGCGGATGCAT 2937

QY 676 ATCTTAAGTTTACGGACTCTCAATGACCACTATCAGTCACTGCTGGCCCAACGTCAG 735
DB 2938 ATCTCAGGTTTACCGGGTGTCTCATGACCACTCAGCACTGGCTGGCCAACTGAA 2997

QY 736 TACCAGCTTAGGAAACGGGGGACAAAATTTCTGAAAACATGACAAAGGCCACCCC 795
DB 2998 TACCAGCTTCGAAGGACAGGTGGAACAAAGTTCTCTCAAAACTTGGACACTGGCCACCCC 3057

QY 796 ATCTTTTATTTGAGTGTGCTCCAGTTCAGAACCCCTCTTACCTACATCACTGCTAC 855
DB 3058 GTCTCTTTTGTAAACGATGTGGTGTCCAAATCAGGACTCTTCCAGGTACATCACTGCTAC 3117

QY 856 TTAGAATCTCACTGGGTTTCCAAATGAAGGACATGACCCGGTTGTGAGTGGACCAAGAA 915
DB 3118 CTAGAGTCACTTAGGCTTCCGGCTAGCGGACTTATCCAAACTGTCCACCGAACAGATT 3177

QY 916 AGCAAGGTGAGCAAGAGATCTCCCGGTTATCTCGGCTCAGAGGTCTCCAGAAACAATA 975
DB 3178 AACA-----GTCAAGTACACAAACCAAGTCAACCGTTCAGAAAAAATGGTGACGTCC 3228

QY 976 GTCCCGAAGAGGACACAGACTCTAAATTCAGTGTAAAGTGTGCTGTGCGGACATTTGTG 1035
DB 3229 TCCCGGAGGAAGAGCTGGGACTTCTTATAGTCAAACTTTGCAATCGGACCTTTGGC 3288

QY 1036 AGCAACATGCGGTAAACTCCACCTAGCAAGCAAAAGCAGCAAGTCAACCCGAAACCAT 1095
DB 3289 AGCAAGCAGCGCTGTAAACTTCCACCTTAGCAAAACACACGGGAAATCTCCGGAAGACCAC 3348

QY 1096 TCACAGTTTGTAAACAGAGCTGGATGAAGATAGC 1129
DB 3349 CTTCTGTATGCTCTGAGTTAGAGAGCAGTAGC 3382

RESULT 14
US-09-739-707-1
; Sequence 1, Application US/09739707
; Patent No. US20020142298A1
; GENERAL INFORMATION:
; APPLICANT: Chen, Hong
; Friemer, Nelson, B.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR
; THE DIAGNOSIS AND TREATMENT OF NEUROPSYCHIATRIC
; DISORDERS
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036/2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/739,707
; FILING DATE: 23-Mar-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/828,008
; FILING DATE: 1997-07-30
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7853-071
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-790-9090
; TELEFAX: 212-869-8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1144 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: Fsh 22
; LOCATION: 1...282
; OTHER INFORMATION:
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-739-707-1

Query Match 19.9%; Score 256.4; DB 9; Length 1144;
Best Local Similarity 67.2%; Pred. No. 2.9e-69;
Matches 437; Conservative 0; Mismatches 206; Indels 7; Gaps 5;

QY 480 GAAGTCGGAATCGATGTCCAGCGCTTTTCAGAGTGTCTCAGTGAAGTCTCAACTTTGCA 539
DB 63 GAAGTCGATGCTGAGCGAGTGTCTTTCAGGAGGCGTTGGACGAGCTGTCCACCGTCCA 122

QY 540 TAAAGAAAAGGCGCGCAGTCCCACTGGAATCTTCAGCATCTTCTGATTTCTACAAGCCCA 599

```
Db 123 CAAAGAGAGGCGCGGCGAGTCCAACTGGAAACCGGAGCA -CTTCTCATCTCGAGGCCCA 181
Qy 600 GTTTCCTCGAGCCTCTTCCAGACATCAGAGGCGAAATACCTGCTGTCTGTCTGTGGGCC 659
Db 182 GTTTCCTCGAGCCTGCGGAGACACAGAGGCGAAGTACATCATGTGCGACTTGGGCC 241
Qy 660 ACAAGAGGATGCAATCTCTAAGTTTACGGGACTCTCAATGAGCACTATCAGTCACTG 719
Db 242 ACAGAGAGGAGTGCATCTCGAAGTTTACTGGGCTCTCCATGACCACTCAGCCACTG 301
Qy 720 GCTGCCCAAGTCAGTACCACTAGGAAACGGCGGAGCAAAATTTCTGAAACAT 779
Db 302 GCTGCCCAATGTGAAGTACCACTGAGGAGACAGGGGAAAGAAATTCCTAAGAACCT 361
Qy 780 GGACAAAGGCCACCCCTCTTTTATGTGAGTACTGTGCTCCCAAGTTTCAAGAACCCCTTC 839
Db 362 GGACACAGGCTATCTGTTTCTTTTGGCAACGATGTGCTCTCAGTTTCAAGTCTGCTTC 421
Qy 840 TACCTACATCAGTCACTTAGAATCTCAGCTGGTTCCTGTTTCAATGAGGACATGACCGCTT 899
Db 422 TACATACATAAGATCTTTTGGAGACACTTGGGCTTTCAGCTGAGGATCTCTCAAGCT 481
Qy 900 GTCAGTGGACAGCAAGCAAGTGGAGCAAGAGATCTCCGGGTATCGTGGCTCAGAG 959
Db 482 GCCACT---CAATCAGATTCAGAAACAGCAGATGTTTGGAAAGTCTTCACCAACAAAC 538
Qy 960 GTCTCCAGAAACAATAGTGTCCGAAAGAGGACACAGACTCTAAATTCAGAGTGTAAAGTTGTG 1019
Db 539 TCTGGGCCCATCGGGGCCACCGAGAGACTTGGGCTCCACTTCCAATGTAAAGCTGTG 598
Qy 1020 CTGTGCGACATTTGAGAGCAACATGCGGTAAACTCCACCTAAGCAAAAGCAGCAAA 1079
Db 599 CAACCGGACTTTTGGAGACAGCAGTCAAA-TGCACCTTTAGTAAGACCCACGCGC-A 655
Qy 1080 GTCACCCCAACACCATTCACAGTTTGTAAACAGCTGATGAGATAGC 1129
Db 656 GTCTCCGAGGACCACTGATCTATGTGACTGAGTTGGAGAAACAGTAGC 705
```

RESULT 15

```
US-10-062-674-1795
; Sequence 1795, Application US/10062674
; Publication No. US2004000559A1
; GENERAL INFORMATION:
; APPLICANT: Loring, Jeanne F.; Kaser, Matthew R.
; TITLE OF INVENTION: MARKERS OF NEURONAL DIFFERENTIATION AND MORPHOGENESIS
; FILE REFERENCE: PA-0026-1 CIP
; CURRENT APPLICATION NUMBER: US/10/062,674
; PRIOR FILING DATE: 2002-01-30
; PRIOR APPLICATION NUMBER: US 09/625,102
; NUMBER OF SEQ ID NOS: 2217
; SOFTWARE: PERL Program
; SEQ ID NO 1795
; LENGTH: 4839
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US2004000559A1 252081.2
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1) ... (4839)
; OTHER INFORMATION: a, t, c, g, or other
```

US-10-062-674-1795

```
Query Match 16.6%; Score 214.2; DB 17; Length 4839;
Best Local Similarity 59.3%; Pred. No. 9.2e-56;
Matches 629; Conservative 0; Mismatches 393; Indels 39; Gaps 14;
Qy 52 AGCAATGGGTGGCGCTCGCCCAACCAAGCCCGGCTTGGCCATGATCAACCCACTCAGC 111
```

```
Db 2408 AACAACTGGGGATCATATGGAACAACCTACCGGAGCCTTCTTTCATCAACCCGCTGAGC 2467
Qy 112 GCCCTGAGTCCGCTCTTGAAACAATCACTTGGGCAAGCCACGAGCCCTTGGCTCACT 171
Db 2468 GCTTTGAGTCCATCATGAAACACCACTCTGGGCAAGGTGTCCAAGCCCGTG----- 2518
Qy 172 TCCTGCTCCAGCCCAAGTTCAAGCACAAATTTCCATGTTCCACAAAGTCAATCTCAATGTC 231
Db 2519 -----AGTCCTCGCTGGACCCGCTGGCGATGCTGTATAAGATCAGCAACAGCATG 2569
Qy 232 ATGACAAAGCCGCTCTTGTAGTCTGCTCCCTCCAC---AAGTCAAGCCAGCGGTGTCCAGGGCC 288
Db 2570 CTGGGACAGACACAGTGTACCCCGCCACCCCTGTGAAGCAGGCGGATGCCATCAGCCG 2629
Qy 289 TACTGTTTGTGAGACA-GCGATCAGCCCATTTGACTGACCAAGTCCAAAGCAAGCAAGCAAG 347
Db 2630 TACTATTATGAAACAGGCGACCAAGCCATTTGACTTAACCAAGTCCAAAGCAAGCCGCT 2689
Qy 348 CGAGTCTCTCGCAAGCACAAATCTTGTATGTCTCCACCT--CAGAACACGCTCTGTCTGACA 406
Db 2690 GGTGTCCAGCGTGTGCTGATTCGGTGGCATCACTTCTGCGGGAGAGCGACTCATGAGACA 2749
Qy 407 TCGCGCATGCTCAAGTCTCTCCCAAGGCAACACCCCAAGCCAGCCCTCTCTCTCCA 466
Db 2750 TCTCGACATGCTGAAACAACTCAGAGCCGCTGACGCCCAAGTCTCTCAGCCCTCCA 2809
Qy 467 GGGTCCCCCATGAAGCTGGAATGTGAGCGCTTTCAGCGCTTTCAGATGTCCTCAGTGAAG 526
Db 2810 CAGTTTCAGA---GAAGTCCGATGCTGATGCGAGCAGCTTTGAGGAGCGTGTGACGAGC 2866
Qy 527 TCTCAACTTTGCATAAAGAAAGAGGCGCGGAGT--CGAACTGGATCC--TCAGCATCTTC 583
Db 2867 TGTCAACCGTTCACAAGAGGAGGCGCGGAGTGTGCCAACTGGAAACCCGTCAGCACTTC 2926
Qy 584 TGATTTTACAAGCCAGTTTTCCTCGAGCTCTTC--CAGACATCAGAGGGCAAAATACCT 641
Db 2927 TCATCTGCGAGCCCGATTTGCTCTGAAAGCTTGGGGAGACCAAGAGGGCAAGTACAT 2986
Qy 642 GCTGTC-TGATCTGGGCCCCACAAGAGCGTATGCAAAATC-TCTAAAGTTTACGGGACTCTCA 699
Db 2987 CATGTGCGGACTTGGGCGCGAGGAGGGTGCATCTTCGAAGTTTACTGGGCTCTCC 3046
Qy 700 A-TGACCACTATCAGTCACTGCTGGTGGCCCAAGCTCAA---GTACCAGCTTAGGAAACGGG 755
Db 3047 AGTGACCACCATCAGCCACTGCTGGCCCAATGTTGAAGTTTACCAGTTTGGAGGAGCAGG 3106
Qy 756 CGGAGCAAAATTTCTGAAACAACTGGCAAAAGGCGCCCATCTTTTATTTGAGTGAAGTCTG 815
Db 3107 GGGAAAGAAATCTTAAGAAACCTGGACACAGGGCATCTGTCTTTCTTTTGAACGATTTG 3166
Qy 816 TGCCTCCCGAGTTCAAGAACCCCTTCTACCTACATCAGTCACTTAGAATCTCACTCGGTTT 875
Db 3167 TGCCTCTCAGTTCAAGACTGCTTCTACATACATAAGTCAATTTGGAGACACACTTGGGCTT 3226
Qy 876 CCAATGAGGACATGACCCGCTTGTGAGTGACAGCAAGCAAGAGGTTGGAGCAAGAT 935
Db 3227 CAGCTTGAAGGATCTCTCCAAGCTGCCACTCAATCAGATTCAAGAACACGAGATGTTTC 3286
Qy 936 CTCGCGGTATCGTGGCTCAGAGTCTCCAGAAACAAATAGTGCAGAGGAGCACACAGA 995
Db 3287 GAAAGTCTTCAACCAACAAACTCTGGGCCCCACTGGGGGCCACCGAGAGAGACTTTGGG 3346
Qy 996 CTCTAAATTCAGAGTGAAG-TTGTGCTGTGCGACATTTTGTGAG-CAAAATCGCGTAAAA 1053
Db 3347 CTCACATACCAAGTAAAGTACTGCAACCGGACTATAGCGAGACAAAGCAGCACTCAA 3406
Qy 1054 CTCACCTTAAGCAAAACGCAAGCAAGTCAACCCGAAACCA 1094
Db 3407 CTGACCTTTAGTAAGACCCAGCGCAAGTCTCCCGAGGACCA 3447
```

Search completed: May 5, 2005, 12:12:34
Job time : 1785.09 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model
Run on: May 4, 2005, 23:17:50 ; Search time 5451.8 Seconds
(without alignments)
8992.766 Million cell updates/sec

Title: US-08-731-499-5
Perfect score: 1288
Sequence: 1 GAGGCACGAGAGAGAA.....AAAAAAAAAAAAAAAAATTC 1288

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:
1: gb_est1:
2: gb_est2:
3: gb_hc:
4: gb_est3:
5: gb_est4:
6: gb_est5:
7: gb_est6:
8: gb_gssi:
9: gb_gss2:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1280.2	99.4	4149	3 HSM806854	BX640770 Homo sapi
2	793.6	61.6	1705	3 AK045693	AK045693 Mus muscu
3	682	53.0	727	1 AI889627	AI889627 wm59g03.x
4	677.4	52.6	680	6 CA449448	CA449448 UI-H-E11-
5	627	48.7	778	6 CB249889	CB249889 UI-M-EXO-
6	613.6	47.6	639	2 AWS16191	AWS16191 xt62d04.x
7	525.8	40.8	892	7 CN538481	CN538481 UI-M-HUO-
8	524.8	40.7	671	7 CN538481	CK435555 909277 MA
9	512.8	39.8	770	5 BU204906	BU204906 604157793
10	505	39.2	953	2 BF140337	BF140337 601787846
11	493	38.3	494	7 CR745711	CR745711 CR745711
12	475.4	36.9	477	1 AA873869	AA873869 oh79d06.8
13	465.2	36.1	686	7 CK773713	CK773713 962388 MA
14	438	34.0	1221	5 BQ715061	BQ715061 AGENCOURT
15	435.2	33.8	581	7 CK775937	CK775937 967078 MA
16	429	33.3	506	4 BM255904	BM255904 518125 MA
17	426	33.1	426	1 AA406466	AA406466 zv10c09.r
18	410.6	31.9	578	4 BG087988	BG087988 H3147C07-
19	406.8	31.6	492	1 AI593964	AI593964 vo99g12.y
20	393.6	30.6	2148	9 AY399842	AY399842 Mus muscu
21	393.6	30.6	3369	3 BC049795	BC049795 Mus muscu
22	393.6	30.6	3370	3 BC049799	BC049799 Mus muscu
23	390.4	30.3	2059	3 BC010679	BC010679 Homo sapi
24	390.4	30.3	2625	3 BC024152	BC024152 Homo sapi

25	389.4	30.2	2127	9 AY399840	AY399840 Homo sapi
26	382.6	29.7	417	7 HI6954	HI6954 ym34b09.81
27	381.4	29.6	1958	3 AK017514	AK017514 Mus muscu
28	379.2	29.4	458	7 HI2950	HI2950 y170h05.r1
29	378.6	29.4	529	2 AW821942	AW821942 uq35h05.y
30	377	29.3	450	7 CK388890	CK388890 L0935G05-
31	376.8	29.3	496	1 AA763786	AA763786 vo99g12.r
32	373	29.0	373	4 BM672622	BM672622 UI-B-CRO-
33	373	29.0	445	7 CK387442	CK387442 L0915G06-
34	369.2	28.7	430	4 BM105530	BM105530 509005 MA
35	369	28.6	514	7 HI6953	HI6953 ym34b09.r1
36	361	28.0	641	5 BP144926	BP144926 BP144926
37	359.6	27.9	505	2 BF193928	BF193928 245665 MA
38	356.8	27.7	486	4 BG070003	BG070003 H3083E07-
39	333.4	25.9	518	7 CN696106	CN696106 E0374D11-
40	331.8	25.8	578	7 CF739142	CF739142 UI-M-HDO-
41	330	25.6	547	5 BU102107	BU102107 PRODIG1D
42	329.8	25.6	841	7 CF242878	CF242878 AGENCOURT
43	327	25.4	488	4 BM288023	BM288023 528879 MA
44	319	24.8	2127	9 AY399841	AY399841 Pan trogl
45	318.8	24.8	953	6 BY717304	BY717304 BY717304

ALIGNMENTS

RESULT 1
HSM806854 4149 bp mRNA linear HTC 22-SEP-2004
LOCUS Homo sapiens mRNA; cDNA DKFZp686F07129 (from clone DKFZp686F07129).
DEFINITION BX640770
ACCESSION BX640770
VERSION BX640770.1 GI:34364881
KEYWORDS HTC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 4149)
AUTHORS Ansoorge, W., Krieger, S., Regiert, T., Rittmueller, C., Schwager, B.,
Mewes, H.W., Weil, B., Amid, C., Osanger, A., Fobo, G., Han, M. and
Wiemann, S.
CONSTRMT The German cDNA Consortium
TITLE Direct Submission
JOURNAL Submitted (22-SEP-2004) MIPS, Ingolstaedter Landstr.1, D-85764
Neuherberg, GERMANY
COMMENT Resequenced by EMBL (European Molecular Biology Laboratories,
Heidelberg/Germany) within the cDNA sequencing consortium of the
German Genome Project.
This clone (DKFZp686F07129) is available at the RZPD Deutsches
Resourcenzentrum fuer Genomforschung GmbH in Berlin, Germany.
Please contact RZPD for ordering:
http://www.rzpd.de/cgi-bin/products/cl.cgi?cloneID=DKFZp686F07129
Further information about the clone and the sequencing project is
available at http://mips.gsf.de/projects/cdna/
Location/Qualifiers
1. 4149
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="RZPD:DKFZp686F07129Q"
/db_xref="taxon:9606"
/clone="DKFZp686F07129"
/tissue type="uterus"
/clone lib="686 (synonym: hlcc3). Vector pSport1_sfi; host
DH10B; sites SfiIA + SfiIB"
/dev_stage="adult"
/note="Chromosome 20 open reading frame 17 protein"
1. 4149
/gene="DKFZp686F07129"
888. -3992
/gene="DKFZp686F07129"
/codon_start=1

/product="hypothetical protein"
/protein_id="CAE45871.1"
/db_xref="GI:34364882"
/db_xref="UniProt/TREMBL:Q6N003"
/translation="MPRRQQAPKRAAGYAEQLEKBEIEKEBEEDSGSVAQLQG
GNDTGDELETPEQKGFCSYQNSPGSLHSDAENESLLSDAQDSDIKSVCGRD
ASDKAHTVSLFNEAHNCMDKMTAVYANILSDSYWSGLGFLKLSNERRNCRTNRG
SNKSDHMDQDALSKSLOQLPSSVSKPSLFSVOLCRQSSKMGCTVFTGASRFR
OCSATVDLVELTVMHNETGHYDDNRKDKLRPTYSKPKRAAFQMDKEDAOKVLK
CMFGSDSPSLDLSVHMLIKHYOKVPLKEPVPTISKWVTPAKRVPDVNRPCSPD
STTGSFADSPSSQKNANLQSSNNRYGONGASYTWQPEACKSOLIKMECSGSDTL
QQLTHMVTYHFLPYSASKKGQVLVDPLAVEKMQSLSEAPNSDSLAPKPSNSA
SDCTATTEKKESEKPEETS KDKVKVSDIEDPLQKLPDPTIKYGLREEDLE
GSKGGDILUKSLNTVTTAINKAQGPSWSPYTHAAYQLSEGTGPPPLPMGSOVLQ
IRPNLTKLRPTAPKVMKPLVSMPTHLAPYQVKKESDKDAVEKCKEKPHEAS
SPSHSGDSFRKSETPPEAKTELGLKEEKLMEKSGSEKPEPLPTALSNGCAL
ANHAPALPCINLUSALQSVLNHLKATPLKSPSCSPSSSTISFMFKSNLNVMDKP
VLSPPSTRSRYLFFENSDDQIDLTKSKKKAESSQAQSCMPPOKHALSDADM
VKVLPKATTPKPASSSRVPMKLEMDRPFVSSEVSTLHKRKGQSNWNPQHLLIL
QAQFASLFTSEGKYLSDLGPOERMQISKFTGLSMVTIISHLANVKYQLRKTGKTK
FPKMDKGHPHIFVCSDAQSOFPTPIYISHLESHLGFOKMDTRLSVDQSKVEQFIS
RVSSAQSPETTAABEDTSPKPKLCCRTFVSKHAVKHLKTHSKSPHHSQFVTD
VDEE"

ORIGIN

Query Match	99.4%;	Score 1280.2;	DB 3;	Length 4149;	
Best Local Similarity	99.8%;	Pred. No. 0;			
Matches 1282;	Conservative	0;	Mismatches	3;	Indels 0; Gaps 0;
Qy	1	GAGGCGAGGAGAGGAGAAACCCAGCCCTGGAGCCACATCTCTCTGAGCAATGGG	60		
Db	2865	GAGGCGAGGAGAGGAGAAACCCAGCCCTGGAGCCACATCTCTCTGAGCAATGGG	2924		
Qy	61	TGCGCCCTGCGCAACACGCGCCCGCCCTGCCATGATCAACCCATCAGCGCCCTGCAG	120		
Db	2925	TGTGCGCCCTGCGCAACACGCGCCCGCCCTGCCATGATCAACCCATCAGCGCCCTGCAG	2984		
Qy	121	TCCGTCTGAAACATCACTTGGGCAAAAGCCAGCGAGCCCTTGGCGCTCACCTCTCTGCTC	180		
Db	2985	TCCGTCTGAAACATCACTTGGGCAAAAGCCAGCGAGCCCTTGGCGCTCACCTCTCTGCTC	3044		
Qy	181	AGCCCAAGTTCAAGCACAATTTCCATGTTCACAAAGTCGAATCTCAATGTGATGGAAGA	240		
Db	3045	AGCCCAAGTTCAAGCACAATTTCCATGTTCACAAAGTCGAATCTCAATGTGATGGAAGA	3104		
Qy	241	CGGCTCTGAGTCTCGCTCCCAAGGTCAGCCAGCGTGTCCAGGCGCTACCTGTTTGG	300		
Db	3105	CGGCTCTGAGTCTCGCTCCCAAGGTCAGCCAGCGTGTCCAGGCGCTACCTGTTTGG	3164		
Qy	301	AACAGCGATCAGCCCATTTGACCTGACCAAGTCCAAAAGCAAGAACCCAGTCTCTCGCAA	360		
Db	3165	AACAGCGATCAGCCCATTTGACCTGACCAAGTCCAAAAGCAAGAACCCAGTCTCTCGCAA	3224		
Qy	361	GCAAAATTTGATGTGTCCTCCACCTCAGAACGCGCTGTGTGATGATGCGCGCATGGTC	420		
Db	3225	GCAAAATTTGATGTGTCCTCCACCTCAGAACGCGCTGTGTGATGATGCGCGCATGGTC	3284		
Qy	421	AAAGTCTCCCAAGCGCCACCCCAAGCGAGCTCTCTCCAGGCGTCCCCCCCATG	480		
Db	3285	AAAGTCTCCCAAGCGCCACCCCAAGCGAGCTCTCTCCAGGCGTCCCCCCCATG	3344		
Qy	481	AACTGGAATGATGTGAGCGCGCTTTGAGGATGTCTCCAGTGAAGTCTCAACTTTGCAT	540		
Db	3345	AACTGGAATGATGTGAGCGCGCTTTGAGGATGTCTCCAGTGAAGTCTCAACTTTGCAT	3404		
Qy	541	AAAAGAAAAGCGCGGAGTCCAACTGGAATCTCAGCATCTCTGATTTACAAAGCCAG	600		
Db	3405	AAAAGAAAAGCGCGGAGTCCAACTGGAATCTCAGCATCTCTGATTTACAAAGCCAG	3464		
Qy	601	TTTGCGCTCGAGCTCTTCCAGACATCAGAGGGCAATACTGCTGTGATCTGGGCGCA	660		
Db	3465	TTTGCGCTCGAGCTCTTCCAGACATCAGAGGGCAATACTGCTGTGATCTGGGCGCA	3524		

Qy	661	CAAGAGCGGTATGCAAAATCTTAAAGTTTACGGGACTCTCAATGACCACTATCAGTCACTGG	720		
Db	3525	CAAGAGCGGTATGCAAAATCTTAAAGTTTACGGGACTCTCAATGACCACTATCAGTCACTGG	3584		
Qy	721	CTGGCCAAAGCTCAAGTACCAAGCTTAGGAAAACGGGGGGGACAAAATTTCTGAAAAACATG	780		
Db	3585	CTGGCCAAAGCTCAAGTACCAAGCTTAGGAAAACGGGGGGGACAAAATTTCTGAAAAACATG	3644		
Qy	781	GACAAAGGCGACCCCATCTTTTATTTCAGTGTACTGTGCTCCAGTTTCAGAAACCCCTTCT	840		
Db	3645	GACAAAGGCGACCCCATCTTTTATTTCAGTGTACTGTGCTCCAGTTTCAGAAACCCCTTCT	3704		
Qy	841	ACCTACATCAGTCACTTAGAAATCTCACCTGGGTTCCTCAAAATGAAGACATGACCCGCTTG	900		
Db	3705	ACCTACATCAGTCACTTAGAAATCTCACCTGGGTTCCTCAAAATGAAGACATGACCCGCTTG	3764		
Qy	901	TCAGTGGACCAAGAACAGTGGAGCAGAGATCTCCCGGGTATCGTCCGCTCAGAGG	960		
Db	3765	TCAGTGGACCAAGAACAGTGGAGCAGAGATCTCCCGGGTATCGTCCGCTCAGAGG	3824		
Qy	961	TCCTCAGAAAACATAGCTGCCGGAAGGACACAGACTCTTAAATTCAGTGTAAAGTTGTGC	1020		
Db	3825	TCCTCAGAAAACATAGCTGCCGGAAGGACACAGACTCTTAAATTCAGTGTAAAGTTGTGC	3884		
Qy	1021	TGTCGGACATTTGTGAGCAAAACATCGGTAAAACTCCACTAAGCAAAAACGACAGCAAG	1080		
Db	3885	TGTCGGACATTTGTGAGCAAAACATCGGTAAAACTCCACTAAGCAAAAACGACAGCAAG	3944		
Qy	1081	TCACCGGAACACCATTCACAGTTTGTACAGACGTGGATGAAGAATAGCTCTGCAGGACG	1140		
Db	3945	TCACCGGAACACCATTCACAGTTTGTACAGACGTGGATGAAGAATAGCTCTGCAGGACG	4004		
Qy	1141	AATGCTTAGTTTCCACTTTCCAGCTGATCCCTCACACTGAACCCCTTCTTCTGTGCA	1200		
Db	4005	AATGCTTAGTTTCCACTTTCCAGCTGATCCCTCACACTGAACCCCTTCTTCTGTGCA	4064		
Qy	1201	CCATCTGCTTCTGACATTTGAATCTTGAATCTCTCTGACACCCCTGCTCTGAGAAGA	1260		
Db	4065	CCATCTGCTTCTGACATTTGAATCTTGAATCTCTCTGACACCCCTGCTCTGAGAAGA	4124		
Qy	1261	CTGCCCAAAAAAAAAAAAAAAAAAAAAA	1285		
Db	4125	CTGCCCAAAAAAAAAAAAAAAAAAAAAA	4149		
RESULT 2					
AK045693		1705 bp	mRNA	linear	HTC 03-APR-2004
LOCUS					
DEFINITION					
ACCESSION					
VERSION					
KEYWORDS					
SOURCE					
ORGANISM					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					

AUTHORS

Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitzunai, T., Tashiro, H., Itoh, M., Suni, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwaki, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Iihikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
20530913
11076861

TITLE

The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.

JOURNAL

Functional annotation of a full-length mouse cDNA collection

MEDLINE

Nature 409, 685-690 (2001)

PUBMED

11076861

REFERENCE

4

AUTHORS

The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.

TITLE

Functional annotation of a full-length mouse cDNA collection

JOURNAL

Nature 409, 685-690 (2001)

REFERENCE

5

AUTHORS

The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.

TITLE

Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs

JOURNAL

Nature 420, 563-573 (2002)

REFERENCE

6 (bases 1 to 1705)

AUTHORS

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hasegawa, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hirooka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.

TITLE

Direct Submission

JOURNAL

Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

COMMENT

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Please visit our web site for further details.

URL: http://genome.gsc.riken.jp/

URL: http://fantom.gsc.riken.jp/

FEATURES

source

1..1705
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM DB:B230304M06"
/db_xref="taxon:10090"
/clone="B230304M06"
/sex="male"
/tissue_type="corpora quadrigemina"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
416..1570
/note="unnamed protein product; TEASHIRT 2 (FRAGMENT) homolog [Mus musculus] (SPTR|Q9JUL7|, evidence: FASTA, 98.6%id, 6.2%length, match=207)
putative)
/codon_start=1
/protein_id="BAC32459.1"
/db_xref="GI:26337547"
/translation="MKSLKQCGKSPHEEATSFQSGESFSKIEPPSPSRKABCPPL

CDS

11076861

KEEKKPOKEPEPELVSSSLTNGCAPANHTPALPSINPLSALQSVLNNHOGKATEPLR
SPSCSSPNSSTSPVHFSSLLHVDKPVISITRPAASVARHYLPENTDQIDLTJTSK
SKRAESSQASQSTPPQKHALCDIADWVKVLPKRAITTPKPAASSRVPKWLKIDVRFE
DVSEVSTLHKRGRQSNWNPQHLILQAFSLFQTSBKGLLSLDGPGRWQISK
FTGLSMTTISHLWLVYQKRTGKFLKNDKMGHPIFYGSCDCAQFRTSTYISHL
ESHLGFMQKDWTRMAADQSKVEQIBSRVSSAQRSPETIAGEEDTDSKFK"

ORIGIN

Query Match 61.6%; Score 793.6; DB 3; Length 1705;
Best Local Similarity 87.4%; Pred. No. 2.4e-206;
Matches 881; Conservative 0; Mismatches 124; Indels 3; Gaps 1;
QY 11 AGAAGGAGAAACCCAGCCCTCGAGGCCACATCTGCTCTCAGCAATGGTGGCGCCCTCG 70
DB 567 AGAAGGAGAAACCCAGGCCCTTAGAACGATATCTTCTGACCAATGGTGTGGCGCGG 626
QY 71 CCAACCAAGCCCGCCCTGCGCATGCAATCAACCACTCAGCGCCCTGCGTCCGTCCTGA 130
DB 627 CTAACCAAGCCCGCCCTGCGCTTCCATCAACCCGCTCAGTGGCCCTGCGTCCCTGA 686
QY 131 ACAATCACTTGGGAAAGCCAGGAGCCCTTGGCGTCACTTCTGCTCCAGCCCAAGTT 190
DB 687 ACAACCAAGCCCGCCCTGCGCATGCAATCAACCACTCAGCGCCCTGCGTCCGTCCTGA 746
QY 191 CAACCAATTTCCATGTTTCCACAGTCCATCTCAATGTCTATGACACAAGCGGCTTGA 250
DB 747 CAAGCAAGCCCGCCCTGTTTCAATCAATCCAGCCCTCCAGTGGTGGACAAGCGGTCTAA 806
QY 251 GTCCTGCTCCACAAAGGTCA---GCCAGCGTGTCCAGCGCTACTCTGTTTGAGAACAGCG 307
DB 807 GTCCACCTCCACAGCGCGCGCGGCGGCGAGACACTACTGTTTGAACACAGCG 866
QY 308 ATCAGCCCAATGACTGACCAAGTCCAAAGCAAGAAAGCGAGTCTCTGCAAGCACAAAT 367
DB 867 ACCAGCCCAATGACTGACCAAGTCCAAAGCAAGAGAGCGAGTCTCTGCAAGCACAAAT 926
QY 368 CTTGTATGTCCCACTCAGAACGAGTCTGTCTGATCTGACATCGCCGACATGTTCAGTCC 427
DB 927 CTTGACGTCCTCCACTCAGAACGATGCTGTGTGTGACATGCTGTATATGTGTGTCAGG 986
QY 428 TCCCAAGCCACCAACCAAGCGAGTCTCTCTCCAGCGGTCCCGCCCATGAAAGTGG 487
DB 987 TCCCAAGCCACCAACCAAGCGAGTCTCTCTGAGGGTCCCTCTATGAGTGG 1046
QY 488 AAATGGATGTCCAGCGCTTTCAGGATGCTCAGTGAAGTCTCAACTTTTGCATAAAGAA 547
DB 1047 AAATGACGTCAGACGCTTCAGGATGTTTCAGGCGAAGTCTCCACTTTTGCACAAAAG 1106
QY 548 AAGCGCGCAGTCCCACTGGAATCTCTGATCTTCTGATTTCTACAAGCCCAAGTTTGGCT 507
DB 1107 AAGCGCGCAGTCCCACTGGAATCTTCTGATTTCTACAAGCCCAAGTTTGGCT 1166
QY 608 CGAGCTCTTCCAGACATCAGAGGCGCAATACCTGCTCTGTGATCTGGGCCCCACAGAGC 667
DB 1167 CGAGCTCTTCCAGACATCAGAGGCGCAATACCTGCTCTGTGATCTGGGCCCCACAGAGC 1226
QY 668 GTATGCAAAATCTTAAGTTTACGGGACTCTCAATGACCACTATCAGTCACTGGTGGCA 727
DB 1227 GGATGCAAAATCTCAAGTTTACCGGACTCTCCATGACCAATCAGCCACTGGTGGCTA 1286
QY 728 AGTCAAGTACAGCTTAGGAAAACGGGCGGGAACAAATTTCTGAAAACATGGACAAAG 787
DB 1287 AGTCAAAATACAGCTTAGGAAAACAGGTGGGACAAAGTCTCTGAAAACATGGACAAAG 1346
QY 788 GCCACCCCATCTTTATTTGAGTGAAGTCTGCTGCTCCAGTCCAGTCCAGTCCCTTCTACCTACA 847
DB 1347 GGCACCCCATCTTTTACTGCGAGTGAAGTCTGCTGCTCCAGTCCAGTCCAGTCCCTTCTACCTACA 1406
QY 848 TCAGTCACTTAGAATCTCACCTGGGTTTCCAAATGAAGGACATGACCCGCTTGTGAGTGG 907
DB 1407 TCAGCCACTAGAGTCTCACCTGGGTTTCCAAATGAAGGACATGACCCGAGTGGCGGTG 1466
QY 908 ACCAGCAAGCAAGGTGAGCAAGAGATCTCTCCCGGTTATCGTTCGGCTCAGAGGTCTCCAG 967

```

Db      1467 ACCAGCAAGAGGTCGAGCAGAGATCTCCGAGTGTGTCGGCTCAGAGGTCTCCAG 1526
QY      968 AAACAATAGCTGCCGAAGAGGACACAGACTCTAAATTCAGTCTAAGT 1015
Db      1527 AAACAATAGCTGCCGAAGAGGACACAGACTCTAAATTCAGTCTAAGT 1574

RESULT 3
AI889627/c
LOCUS   AI889627      727 bp      mRNA      linear      EST 07-MAR-2000
DEFINITION
wm59g03.x1 NCI CGAP Ut2 Homo sapiens cDNA clone IMAGE:2440276 3'
similar to TR:060534 O60534 ANTIGEN NY-CO-33. ;, mRNA sequence.
ACCESSION
VERSION  AI889627
KEYWORDS  EST.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 727)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TUMOR     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
TITLE     Tumor Gene Index
JOURNAL   Unpublished (1997)
COMMENT   Contact: Robert Strausberg, Ph.D.
          Email: cgabbs-r@mail.nih.gov
          Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
          Emmert-Buck, M.D., Ph.D.
          CDNA Library Preparation: Life Technologies, Inc.
          CDNA Library Arrayed by: Greg Lennon, Ph.D.
          DNA Sequencing by: Washington University Genome Sequencing Center
          Clone distribution: NCI-CGAP clone distribution information can be
          found through the I.M.A.G.E. Consortium/LLNL at:
          www-bio.llnl.gov/bbrp/image/image.html
          Insert length: 1472 Std Error: 0.00
          Seq primer: -40UP from Gibco
          High quality sequence stop: 412.
          Location/Qualifiers
            1..727
              /organism="Homo sapiens"
              /mol_type="mRNA"
              /db_xref="taxon:9606"
              /clone="IMAGE:2440276"
              /tissue_type="moderately-differentiated endometrial
              adenocarcinoma, 3 pooled tumors"
              /lab_host="DH10B"
              /clone_lib="NCI-CGAP_Ut2"
              /notes="Organ: uterus; Vector: pCMV-SPORT6; Site 1: SalI;
              Site 2: NotI; Cloned unidirectionally. Primer: Oligo dt.
              Average insert size 1.85 kb. Life Technologies catalog #:
              11539-012"

ORIGIN
Query Match      53.0%; Score 682; DB 1; Length 727;
Best Local Similarity 97.7%; Pred. No. 8.7e-176;
Matches 710; Conservative 0; Mismatches 15; Indels 2; Gaps 2;

QY      544 AGAAAGGCGGCGAGTCCTCACTGGATCTTCAGCA-TCTTCTGATTTCTAC-AAGCCAGT 601
Db      727 AGAAAGGCGGCGAGTCCTCACTGGATCTTCAGCA-TCTTCTGATTTCTAC-AAGCCAGT 668
QY      602 TTGCTCTCGAGCTCTTCCGACATCAGAGGCAATACCTGCTGTCGATCTGGGCCAC 661
Db      667 TGGCTCTCGAGCTCTTCCGACATCAGAGGCAATACCTGCTGTCGATCTGGGCCAC 608
QY      662 AAGAGCGTATGCAATCTCTAAGTTTACGGGACTCTCAATGACCACTATCAGTCTGGC 721
Db      607 AAGAGCGTATGCAATCTCTAAGTTTACGGGACTCTCAATGACCACTATCAGTCTGGC 548
QY      722 TGGCCAAAGTCAGTACAGTCTAGGAAACCGGCGGCGGCAAAATTTCTGAAAAACATGG 781
Db      547 TGGCCAAAGTCAGTACAGTCTAGGAAACCGGCGGCGGCAAAATTTCTGAAAAACATGG 488

```

```

QY      782 ACAAGGCGCACCCCATCTTTTATTGTCAGTGAATGTCCTCCAGTTTCAGAACCCCTTCTA 841
Db      487 ACAAGGCGCACCCCATCTTTTATTGTCAGTGAATGTCCTCCAGTTTCAGAACCCCTTCTA 428
QY      842 CCTACATAGTCACATTAGAAATCTCACTGGGTTTCCAAATGAAGGACATGACCCGTTGT 901
Db      427 CCTACATAGTCACATTAGAAATCTCACTGGGTTTCCAAATGAAGGACATGACCCGTTGT 368
QY      902 CAGTGACACAGCAAAAGCAGGTGGAGCAGAGATCTCCCGGTATCGTCGGCTCAGAGT 961
Db      367 CAGTGACACAGCAAAAGCAGGTGGAGCAGAGATCTCCCGGTATCGTCGGCTCAGAGT 308
QY      962 CTCGAAAAACAATAGCTGCCGAAGAGGACACAGACTCTAAATTCAGTCTAAGTTTGTCT 1021
Db      307 CTCGAAAAACAATAGCTGCCGAAGAGGACACAGACTCTAAATTCAGTCTAAGTTTGTCT 248
QY      1022 GTCCGACATTTGTGAGCAAAACAATGCGGTAAAACTCCACCTAAGCAAAACGACAGCAAGT 1081
Db      247 GTCCGACATTTGTGAGCAAAACAATGCGGTAAAACTCCACCTAAGCAAAACGACAGCAAGT 188
QY      1082 CACCCGACACACATTCACAGTTTGTAAACAGAGCTGGATGAAGATAGCTCTGCAGACCA 1141
Db      187 CACCCGACACACATTCACAGTTTGTAAACAGAGCTGGATGAAGATAGCTCTGCAGACCA 128
QY      1142 ATGCTCTAGTTTCCACTTTCCAGCTTGGATCCCTCAGACTGAACCTTCTTCTGTTGCAC 1201
Db      127 ATGCTCTAGTTTCCACTTTCCAGCTTGGATCCCTCAGACTGAACCTTCTTCTGTTGCAC 68
QY      1202 CATCTGCTTCTGACATGAATCAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1261
Db      67 CATCTGCTTCTGACATGAATCAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 8
QY      1262 TGCCTAAA 1268
Db      7 TGCCTAAA 1

RESULT 4
LOCUS   CA449448/c
DEFINITION
UI-H-E11-ayt-h-09-0-UI.s1 NCI CGAP E11 Homo sapiens cDNA clone
UI-H-E11-ayt-h-09-0-UI 3', mRNA sequence.
ACCESSION
VERSION  CA449448.1 GI:24813868
KEYWORDS  EST.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 680)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS   National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
TITLE     Tumor Gene Index
JOURNAL   Unpublished (1997)
COMMENT   Contact: Robert Strausberg, Ph.D.
          Email: cgabbs-r@mail.nih.gov
          Tissue Procurement: Dr. Jose Mercuende
          CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
          CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
          DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
          Clone Distribution: Clone distribution information can be obtained
          from Dr. M. Bento Soares, bento-soares@uiowa.edu
          Seq primer: M13 FORWARD
          POLYA=Yes.
          Location/Qualifiers
            1..680
              /organism="Homo sapiens"
              /mol_type="mRNA"
              /db_xref="taxon:9606"
              /clone="UI-H-E11-ayt-h-09-0-UI"
              /tissue_type="Chondrosarcoma"
              /dev_stage="Adult"

```

```
/lab_host="DH10B (Life Technologies)"
/clone_lib="NCI_CGAP_E11"
/notes="Organ: Left Pelvis; Vector: pT7T3-Pac (Pharmacia)
with a modified polylinker; Site 1: EcoR I; Site 2: Not I;
NCI CGAP E11 is a normalized cDNA library containing the
following tissue(s): Chondrosarcoma. The library was
constructed according to Bonaldo, Lennon and Soares,
Genome Research, 6:791-806, 1996. First strand cDNA
synthesis was primed with an oligo-dT primer containing a
Not I site. Double stranded cDNA was ligated to an EcoR I
adaptor, digested with Not I, and cloned directionally
into pT7T3-Pac vector. The oligonucleotide used to prime
the synthesis of first-strand cDNA contains a library tag
sequence that is located between the Not I site and the
(dT)18 tail. The sequence tag for this library is
ACACTTGCAC.
TAG_TISSUE=chondrosarcoma
TAG_LIB=UI-H-E11
TAG_SEQ=ACACTTGCAC"

ORIGIN
Query Match 52.6%; Score 677.4; DB 6; Length 680;
Best Local Similarity 99.7%; Pred. No. 1.6e-174;
Matches 678; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 604 GCCTGAGCCTTTCAGACATCAGAGGCAATACCTGCTCTGATCTGGGCCCAAA 663
DB 680 GCCTGAGCCTTTCAGACATCAGAGGCAATACCTGCTCTGATCTGGGCCCAAA 621

QY 664 GAGCGTATGCAATCTCTAAGTTTACGGGACTCTCAATGACCACCTATCAGTGGCTG 723
DB 620 GAGCGTATGCAATCTCTAAGTTTACGGGACTCTCAATGACCACCTATCAGTGGCTG 561

QY 724 GCCAAGCTCAAGTACCAGCTTAGGAAAACGGCGGACAAAATTTCTGAAAAACATGGAC 783
DB 560 GCCAAGCTCAAGTACCAGCTTAGGAAAACGGCGGACAAAATTTCTGAAAAACATGGAC 501

QY 784 AAAGGCCACCCCATCTTTTATTCAGTGTACTGTGCTCCAGTTCCAGACCCCTTCTACC 843
DB 500 AAAGGCCACCCCATCTTTTATTCAGTGTACTGTGCTCCAGTTCCAGACCCCTTCTACC 441

QY 844 TACATCAGTCACTAGAACTCCTACCTGGTTTCCAAATGAAGACATGACCCGCTGTCA 903
DB 440 TACATCAGTCACTAGAACTCCTACCTGGTTTCCAAATGAAGACATGACCCGCTGTCA 381

QY 904 GTGGACCAAGCAAGGTGAGCAAGAGATCTCCCGGGTATCGTCGGCTCAGAGGCTCT 963
DB 380 GTGGACCAAGCAAGGTGAGCAAGAGATCTCCCGGGTATCGTCGGCTCAGAGGCTCT 321

QY 964 CCAGAAACAATAGTGTCCGAAGAGGACACAGACTCTAAATTCAGTGTAAAGTTGTCTGT 1023
DB 320 CCAGAAACAATAGTGTCCGAAGAGGACACAGACTCTAAATTCAGTGTAAAGTTGTCTGT 261

QY 1024 CGGACATTTGTGAGCAAAATCGGTGTAATCTCCACCTAAGCAAAACGACAGCAAGTCA 1083
DB 260 CGGACATTTGTGAGCAAAATCGGTGTAATCTCCACCTAAGCAAAACGACAGCAAGTCA 201

QY 1084 CCGGAACACCATTCACAGTTTGTAAACAGACGTGGATGAAGATAGCTCTGCAGGACGAAT 1143
DB 200 CCGGAACACCATTCACAGTTTGTAAACAGACGTGGATGAAGATAGCTCTGCAGGACGAAT 141

QY 1144 GCCTTAGTTTCCACTTTCAGAGCTGATCCCTTCCACTGAAACCCCTTCTTCGTGCAACA 1203
DB 140 GCCTTAGTTTCCACTTTCAGAGCTGATCCCTTCCACTGAAACCCCTTCTTCGTGCAACA 81

QY 1204 TCGTGTCTTGACATTTGAATCTATTTGAATCTCTCTGACACCCCTGCTCTGAGAGACTG 1263
DB 80 TCGTGTCTTGACATTTGAATCTATTTGAATCTCTCTGACACCCCTGCTCTGAGAGACTG 21

QY 1264 CCAAAAAAAAAAAAAAAAAA 1283
DB 20 CCAAAAAAAAAAAAAAAAAA 1
```

```
RESULT 5
CB249889
LOCUS
DEFINITION
CB249889 778 bp mRNA linear EST 15-JUL-2003
IMAGE:5720974 5', mRNA sequence.
CB249889
ACCESSION
VERSION 1 GI:28389782
KEYWORDS
SOURCE
ORGANISM
Mus musculus (house mouse)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 778)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Dr. James Lin, University of Iowa
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
This clone was contributed by the Brain Molecular Anatomy Project
(BMAP)
Seq primer: pYX-5.
Location/Qualifiers
1. 778
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:5720974"
/tissue_type="whole brain"
/dev_stages="embryo 15.5 dpc"
/lab_host="DH10B (T1 phage resistant)"
/clone_lib="NIH BMAP EXO"
/notes="Organ: brain; Vector: pYX-Asc; Site 1: EcoR I;
Site 2: Not I; The library was constructed according to
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
1996. Denatured mRNA was size fractionated on a 1% agarose
gel. First strand cDNA synthesis was primed with an
oligo-dT primer containing a Not I site. Double stranded
cDNA was size selected according to mRNA size fraction.
ligated with EcoR I adaptor, digested with Not I, and then
cloned directionally into pYX-Asc vector. The library tag
sequence located between the Not I site and the polyA
tail, is GTCGCTGGAA. This library was created for the
University of Iowa Mouse Brain Molecular Anatomy Project
(BMAP). 'Gene Discovery in the Developing Mouse Nervous
System', supported by National Institutes of Mental Health
(NIMH), Hemin Chin, Ph.D., program coordinator."
```

```
Query Match 48.7%; Score 627; DB 6; Length 778;
Best Local Similarity 88.9%; Pred. No. 1.1e-160;
Matches 675; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

QY 257 CTTCCACAAGGTTCAGCCGCGTGTCCAGGCGCTACTGTTTGAGAACAGCGATCAGCCCA 316
DB 8 CCACACGAGCGCGCGCCAGCGTGGCCAGACACTACTGTTTGAGAACACCGACAGCCCA 67

QY 317 TTGACCTCAGCAAGTCCAAAGCAAGACGCGAGTCTCTCGCAAGACACATCTTGATGT 376
DB 68 TTGACCTCAGCAAGTCCAAAGCAAGACGCGAGTCTCTCGCAAGACACATCTTGACCT 127

QY 377 CCCCACCTCAGAACGACGCTCTGTGACATCGCGACATGCTCAAGTCTCCCAAG 436
DB 128 CCCCACCTCAGAACGACGCTCTGTGACATGCTCAAGTCTCTCCCAAG 187
```


REFERENCE 1 (bases 1 to 892)
AUTHORS Smith,T.P.L., Freking,B.A., Ford,J.J., Vallet,J.L., Wise,T.A.,
Nonneman,D.J., Wray,J.E. and Keele,J.W.
TITLE Porcine EST collection using a normalized library constructed from
embryos representing early developmental stages
JOURNAL Unpublished (2003)
COMMENT Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@mail.marc.usda.gov
Single pass sequencing. Bases called with phred v0.020425.c and
trimmed with the aid of the trim_alt option. Vector identified with
cross_match v0.990329.
Plate: TMW8025 row: A column: 10
Seq primer: GTAATACGACTCAGCTATAGG.
FEATURES
Location/Qualifiers
source
1..892
/organism="Sus scrofa"
/mol_type="mRNA"
/db_xref="taxon:9823"
/tissue_type="pooled"
/lab_host="DH10B"
/clone_lib="MARC 4PIG"
/note="Vector: pcDNA3.1; Site 1: EcoRI; Site 2: NotI;
Library made with combined RNA from day-10, day-13,
day-15, day-25, and day-30 whole embryos."
ORIGIN
Query Match 40.8%; Score 525.8; DB 7; Length 892;
Best Local Similarity 90.8%; Pred. No. 7.3e-133;
Matches 560; Conservative 0; Mismatches 57; Indels 0; Gaps 0;
QY 521 GTGAAGTCTCAACTTTCATATAAAGAAAGGCGGCGAGTCCAACTGGAAATCCTCAGCATC 580
Db 1 GCGAGTCTCGACCTTCATATAAAGAAAGGCGGCGAGTCCAACTGGAAATCCTCAGCATC 60
QY 581 TTCTGATCTCAAGCCAGCTTGTGCTCGAGCTCTTCCAGACATCAGAGGGCAATACC 640
Db 61 TTCTGATCTCGAGCTCAGTTTGTGCTCGAGCTCTTCCAGACCTCGAGGGCAATACC 120
QY 641 TGCTGTCTGATCTGGGCGCAGAGCGGTATGCAATCTCTAAGTTTACGGACTCTCAA 700
Db 121 TGCTGTCTGACCTGGGTCCCGAGGAGCAATGCAGATCTCGAAGTTTCAGGACTCTCAA 180
QY 701 TGACCACTATCAGTCACTGGCTGGCCACGCTCAAGTACCAGCTTAGGAAAGCGGCGGGA 760
Db 181 TGACCACTATCAGCCTAGCCATGGCTGGCCACGCTCAAGTACCAGCTTAGGAAAGCGGCGGGA 240
QY 761 CAAATTTCTGAAAAACATGGAACAAAGGCCACCCCATCTTTTATTCAGTCACTGTGCT 820
Db 241 CGAATTTCTGAAAAACATGGAACAAAGGACACCTTATCTTTACTGAGTCACTGTGCT 300
QY 821 CCCAGTTGAGAACCCCTTCTACCTACATCAGTCACTTAGAATCTCACTGGGTTCCAAA 880
Db 301 CCCAGTTCCGAGCGCTTTCAGCTACATCAGCACTTAGAATCTCACTAGGCTTCCAAA 360
QY 881 TGAAGGATGATGACCGCTTGTTCAGTGGCCAGCAAGCAAGTGGAGCAAGAGATCTGCC 940
Db 361 TGAAGGATGATGACCGCTTGTTCAGTGGCCAGCAAGCAAGTGGAGCAAGAGATCTGCC 420
QY 941 GGGTATCGTCTGAGGCTTCCAGAGGCTCCAGAAACAAATAGCTGCCAGAGGACACAGACTCTA 1000
Db 421 GGGTATCGTCTGAGGCTTCCAGAGGCTCAGCGAAACCATAGCTGGGAGGACACAGGACTCTA 480
QY 1001 AATTCAGTGTAAAGTTGTCTGCGGACATTTGTGAGCAAAACATCGCGGTAAACCTCCACC 1060
Db 481 AATTCAGTGTAAAGTTGTCTGCGGACATTTGTGAGCAAAACATCGAGTAAACCTCCACC 540
QY 1061 TAAGCAAAACGACAGCAAGTCAACCGGACCACTTACAGTTTGTACAGACGTTGGATG 1120
Db 541 TAAGCAAAACGACAGCAAGTCAACCGGACCACTTGTGTAACAGACGTTGGATG 600

QY 1121 AAGAAATAGCTCTGCAGG 1137
Db 601 AGGAATRACTCTGCAGG 617

RESULT 8
CNS38481 671 bp mRNA linear EST 29-APR-2004
LOCUS UI-M-HUO-cqs-i-15-0-UI.r1 NIH_BMAP_HUO Mus musculus cDNA clone
DEFINITION IMAGE:30667886 5', mRNA sequence.
ACCESSION CNS38481
VERSION CNS38481.1 GI:46866637
KEYWORDS Mus musculus (house mouse)
SOURCE EST.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 671)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Dr. James Lin University of Iowa
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Distribution information can be found at
<http://genome.uiowa.edu/distribution/mouseefl.html>
This clone was contributed by the Brain Molecular Anatomy Project
(BMAP)
Seq primer: PYX-5.
FEATURES
Location/Qualifiers
source
1..671
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:30667886"
/tissue_type="whole eye"
/dev_stages="newborn (1, 5, 15 days)"
/lab_host="DH10B (T1 phage resistant)"
/clone_lib="NIH_BMAP_HUO"
/note="Organ: Eye; Vector: pYX-Asc; Site 1: EcoR I;
Site 2: Not I; The library was constructed according
Bonaldo, Lennon and Soares, Genome Research, 6:791-806.
1996. Denatured RNA was size fractionated on a 1% agarose
gel. First strand cDNA synthesis was primed with oligo-dT
primer containing a Not I site. Double strand cDNA was
size selected according to mRNA size fraction, ligated
with EcoR I adaptor, digested with NotI and then cloned
directionally into pYX-Asc vector. The library tag
sequence located between the Not I site and the polyA tail
is AATAATACG. This library was created for the University
Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
Developing Mouse Nervous System', supported by National
Institute of Mental Health (NIMH)."
ORIGIN
Query Match 40.7%; Score 524.8; DB 7; Length 671;
Best Local Similarity 91.4%; Pred. No. 1.3e-132;
Matches 556; Conservative 0; Mismatches 52; Indels 0; Gaps 0;
QY 530 CAACCTTCATAAAGAAAGCGCGGAGTCCAACTGGAATCCTCAGCATCTTCTGATTC 589
Db 1 CCACCTTCGCAAAAGGAAAGCGCGGAGTCCAACTGGAACCCCGAGCATCTTCTATCC 60
QY 590 TACAGCCCGAGTTTGCCTCGAGCTCTTCCAGACATCAGAGGGCAATACCTGCTGTCTG 649
Db 61 TCGAAGCTCAGTTTGCCTCGAGCTCTTCCAGACCTCAGAGGGCAATACCTGCTTCTG 120
QY 650 ATCTGGGCGCCAAAGAGCGGTATGCAAACTCTTAAGTTTACGGGACTCTCAATGACCACTA 709

Db 121 ACTGGGCCCCAAGAGCGGATGCAAAATCTCAAGTTTACCGGACTCTCCATGACCACAA 180
Qy 710 TCAGTCACCTGGCTGGCCACAGTCAAGTACAGCTTAGGAAACGGCGGACAAATTTTC 769
Db 181 TCAGCCACTGGCTGGCTAAAGTCAATACCAAGCTTAGGAAACAGGTGGGACAAAGTTCC 240
Qy 770 TGAAGAAACATGACAAAGGCCACCCCATCTTTTATTGACAGTGACTGTGCTCCAGTTCA 829
Db 241 TGAAGAAACATGACAAAGGCCACCCCATCTTTTATTGACAGTGACTGTGCTCCAGTTCA 300
Qy 830 GAACCCCTTCTACCTACATCAGTCACTTAGAATCTCACCTGGGTTTCCAAATGAAGGACA 889
Db 301 GAACCCCTTCTACCTACATCAGCACCCTAGAGTCTCACCTGGGTTTCCAAATGAAGGACA 360
Qy 890 TGACCCGCTTGTGACGTGGACCAAGCAAGGTCGAGCAAGAGATCTCCCGGTATCGT 949
Db 361 TGACCCGCTTGTGACGTGGACCAAGCAAGGTCGAGCAAGAGATCTCCCGGTATCGT 420
Qy 950 CGGCTCAGAGGTTCTCCAGAAACAAATAGCTGCCGAGGACACAGACTCTAAATTCAGT 1009
Db 421 CGGCTCAGAGGTTCTCCAGAAACAAATAGCTGCCGAGGACACAGACTCTAAATTCAGT 480
Qy 1010 GTAAGTTGTGCTGGCAATTTGTGAGCAAAACATGCGGTAAATTCACCTAAAGCAAAA 1069
Db 481 GTAAGTTGTGCTGGCAATTTGTGAGCAAAACATGCGGTAAATTCACCTAAAGCAAAA 540
Qy 1070 CGCAGCAGGTCACCCGACACCATTCACAGTTTGTACAGACGTTGATGAAGATAGC 1129
Db 541 CGCAGCAGGTCACCCGACACCATTCACAGTTTGTAGCAGCGTGGATGAAGATAGT 600
Qy 1130 TCTGCAGG 1137
Db 601 TCTGCAGG 608

RESULT 9
LOCUS BU204906 770 bp mRNA linear EST 25-NOV-2002
DEFINITION 604157793F1 CSEQCHN03 Gallus gallus cDNA clone ChEST1008j24 5',
mRNA sequence.
ACCESSION BU204906
VERSION BU204906.1 GI:25369490
KEYWORDS Gallus gallus (chicken)
SOURCE Gallus gallus
ORGANISM Gallus gallus
REFERENCE 1 (bases 1 to 770)
AUTHORS Boardman, P. E., Sanz-Ezquerro, J., Overton, I. M., Burt, D. W., Bosch, E.,
Pong, W. T., Tickle, C., Brown, W. R. A., Wilson, S. A. and Hubbard, S. J.
TITLE A Comprehensive Collection of Chicken cDNAs
JOURNAL Curr. Biol. 12 (22), 1965-1969 (2002)
MEDLINE 22335534
PUBMED 12445392
COMMENT Contact: Simon Hubbard
Department of Biomolecular Sciences
University of Manchester Institute of Science and Technology
(UMIST)
PO Box 88, Manchester, M60 1QD, UK
Tel: 01612008930
Fax: 01612360409
Email: Simon.Hubbard@umist.ac.uk.
Location/Qualifiers
1. .770
/organism="Gallus gallus"
/mol_type="mRNA"
/strain="White Leghorn, HiseX"
/db_xref="taxon:9031"
/clone="ChEST1008j24"
/tissue_type="whole embryo"
/dev_stage="20-21"

FEATURES
source

/lab_host="DH10B"
/clone_lib="CSEQCHN03"
/note="Organ: whole embryo; Vector: pBluescript II KS(+);
Site_1: EcoRI; Site_2: NotI; This normalized library was
constructed from 1 million independent clones. cDNA
synthesis was initiated using an oligo(dT) primer, using
methylated C in the first strand synthesis reaction.
Following this first strand reaction, double-stranded cDNA
was blunted, ligated to NotI adapters, digested with
EcoRI, size-selected, and cloned into the NotI and EcoRI
compatible sites of a custom modified MCS of the
pBluescript (KS+) vector. The library was normalized in 2
rounds using conditions adapted from Soares et al., PNAS
(1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6
(1996): 791, except that a significantly longer
reannealing hybridization was used."

ORIGIN

Query Match	Best Local Similarity	Score	512.8;	DB 5;	Length	770;			
Matches	616;	Conservative	0;	Mismatches	112;	Indels	6;	Gaps	3;
Qy	419	TCAAAGTCTCTCCCAAGGCCACACCCCAAGCCAGCCCTCTCTCCAGGGTCCCCCA	478						
Db	1	TCAAAGTTCTTCCCAAGGCTACTACACCAAAACCTGCCGCATCTTCAAGGATCCCATCTA	60						
Qy	479	TGAAGCTGGAATGATGTGAGCGCTTTGAGCATGTCTCCAGTGAAGTCTCAACTTTGC	538						
Db	61	TGAAGTTGGAATAGATGTCCGACGCTTTGAGATGTCTCAACAGAGTCTCCACTCTGC	120						
Qy	539	ATAAAGAAAGGCCCGGACGTCCAACTGGAATCTCTCAGCATCTTCTGATTCTACAAGCCC	598						
Db	121	ATAAAGGAAGGCCAGGCGAGTCAAACTGGAACCCCTCAGCATCTTCTGATTCTGCAAGCTC	180						
Qy	599	AGTTTGCCTCGAGCCTTTCAGACATCAGAGGGCAATACCTGCTGTCTGATCTGGGCC	658						
Db	181	AGTGTGCTTCCAGCCTTTCAGACATCTGGAAGTAAATATTTATATCAGATCTAGAGCC	240						
Qy	659	CACAAGAGGTATGAAATCTCTAAGTTTACGGGACTCTCAATGACCATCTACAGTCACT	718						
Db	241	CACAGAGGATGATGAGATCTCAAAATTTTACCTGAGTGTCAATGACCATCTAGCCACT	300						
Qy	719	GGCTGCCAAACGCTCAAGTACCAGTCTAGGAAACCGGGCGGACAAATTTCTGAAAAACA	778						
Db	301	GGTTGGCAATGTCAAGTACCAACTTAGGAAACTGGAGGAAACAAAGTTTTTGAAAAACA	360						
Qy	779	TGACAAAGGCCACCCCATCTTTTATTGACGTGACTGTGCTCCGAGTTTCAGAACCCCTT	838						
Db	361	TGACAAAGGGCATCGGTCCTTTTATTGACGTGACTGTGCTCAGTCTCAGTTTCGAAACCCAT	420						
Qy	839	CTACCTACATCAGTCACTTAGAATCTCACCTGGGTTTCCAAATGAAGGACATGACCCGCT	898						
Db	421	CTACTTACATTAGCCATCTAGAAATCTCATCTAGTTTCCAAATGAAGACATGAACAGGC	480						
Qy	899	TGTCAGTGGACCAAGCAAGGTTGAGCAAGAGATCTCCCGGGTATCGTGGGCTCAGA	958						
Db	481	TGCTGTGGAGCAGCAAAACCAAGGTAGACCAAGAAATCTCCAGAT---TTCAAGTTCAA	537						
Qy	959	GGTCTCCAGAAACAATAGCTGCCGAGAGGACACAGACTCTAAATTCAGTGTAGTTGT	1018						
Db	538	GGTCTCTGAAACAAATAGCTGGAGAGGACACAGACTCTAAGTTCAAATGTAAGTGT	597						
Qy	1019	GCTGTCCGACATTTGTGAGCAAAACATGCGGTAAAACTCCACCTTAAGCAAAACCCACAGCA	1078						
Db	598	GCTGTCCGACATTTGCAAGCAAAACATGCTGTAAACTTCATCTAAGCAAAACACACAGCA	657						
Qy	1079	AGTCACCCG-AAACACCATTCACAGTTTGTAAACAGACGTGGATGAAGATAGCTCTCAGG	1137						
Db	658	AGTCACCAAGAACACCATTCACAAATTTGTAGCAGAGTGGATGAAGATAAATCTAG--GG	715						
Qy	1138	ACGAATGCCCTTAGT 1151							
Db	716	ACGAATGCCCTTAGT 729							


```
RESULT 10
BF140337
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

Mus musculus (house mouse)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 953)
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLM9261 row: 1 column: 12
High quality sequence stop: 676.
Location/Qualifiers
1..953
/organism="Mus musculus"
/mol_type="mRNA"
/strain="CZECH II"
/db_xref="taxon:10090"
/clone="IMAGE:4015643"
/tissue_type="tumor, metastatic to mammary"
/lab_host="DH10B"
/clone_lib="NCI CGAP Lu30"
/note="Organ: lung; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI; transgenic model WNT-1, expression driven by
MTV-LTR enhancer; Cloned unidirectionally. Primer: Oligo
dT. Library constructed by Life Technologies.
Investigator providing samples: Gilbert Smith, NIH"

ORIGIN
Query Match 39.2%; Score 505; DB 2; Length 953;
Best Local Similarity 88.3%; Pred. No. 3.8e-127;
Matches 628; Conservative 0; Mismatches 75; Indels 8; Gaps 7;

QY 418 GTCAAGTCTCCCAAGCCACCAACCCCAAGCCAGCTCTCTCCAGGGTCCCGCC 477
DB 1 GTCAAGTCTCTCCCAAGCCACCAACCCCAAGCCAGCGCTCTCTCGAGGTCCTCT 60
QY 478 ATGAAGTGGAAATCGATCGAGCGCTTTGAGGATGTCTCAGTGAAGTCTCAACTTTG 537
DB 61 ATGAAGTGGAAATCGATCGAGCGCTTTGAGGATGTTCGAGCGAAGTCTCCACTTTG 120
QY 538 CATATAAA-GAAAGCGCGGAGTCGAAATCTTCAGTATCTTCGATCTTCAACG 596
DB 121 CACAAAAGGAAGAGCGCGGAGTCCAACTGGAAACCCCGAGCATCTTCTCATCTGCAAGC 180
QY 597 CCAGTTTGCTCGAGCTCTCCAGACATCAGAGCGCAATACCTGCTGTCTGATCTGGG 656
DB 181 TCAG-TTGCTCGAGCTCTCCAGACCTCAGAGGCAATACCTGCTTCTGACTGGG 239
QY 657 CCCACAAGAGCGTATGCAATCTCTAAGTTTACGGGACTCTCAATGACCACTATCAGTCA 716
DB 240 CCCCCAAGAGCGATGCAATCTCAAGTTTACCAGGACTCTCCATGACCAATCAGCCA 299
QY 717 CTGGCTGGCCACGCTCAAGTACAGCTTAGGAAACGGCGGAGCAAAATTTCTGAAAAA 776
DB 300 CTGGCTGGCTAACGTCATAATACAGCTTAGGAAACAGGTTGGGACAAAGTCTCTGAAAAA 359
```

```
QY 777 CATGGACAAGGCCACCCCATCTTTTATTGAGTGAAGTGTGCTCCAGTTTCAGAACCCC 836
DB 360 CATGGACAAGGGGACCCCATCTTTTACTGAGTGAAGTGTGCTCCAGTTTCAGAACCCC 419
QY 837 TTCTACCTTACATCAGTCACTTAGAATCTCAGCTGGTTTCCAAATGAAGGACATGACCCG 896
DB 420 CTCTACCTTACATCAGCACCTAGAGTCTCAGCTGGGTTCCAAATGAAGGACATGACCCG 479
QY 897 CTCTGAGTGGACCAAGCAAGGTGGAGCAAGAGATCTCCCGGGPATCGTGGCTCA 956
DB 480 GATGGCGGCTCA-CAGCAAGCAAGGTGGAGCAGAGATCTCCCGAGTGTGCTGGCTCA 538
QY 957 GAGTCTCCAGAAACAATAGTCTGCGAAGAGGACA-CAGACTCT-AAATTCAGTGTAAAG 1014
DB 539 GAGTCTCCAGAAACAATAGTCTGCGAAGAGGACACCAAGACTCTAAATTCAGTGTAAAG 598
QY 1015 TTGTGCTGTGCGACA-TTTGTGAGCAACATGCGGTAAACTCCACCTTAAGCAAAACGCA 1073
DB 599 TTATGCCGTGCGACATTTTGTGAGCAACACGAGTAGTAAACTTACCTTAGCAAAACGCG 658
QY 1074 CAGCAAGTCAACCCGAAACCACTTACAGTTTGTAAACAGCGTGGATGAAGA 1124
DB 659 ACAGAGTCAACCCGAAACCACTTCTCAG--TGTAGCAGACGGGATGAAGA 707

RESULT 11
CR745711
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

CR745711 494 bp mRNA linear EST 30-AUG-2004
CR745711 NCI CGAP_Brn23 Homo sapiens cDNA clone IMAGE971N1484 ;
IMAGE:2010852 5', mRNA sequence.
CR745711
CR745711.1 GI:51654109
EST.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 494)
Ebert, L., Heil, O., Hennig, S., Korn, B., Neubert, P., Partsch, E.,
Peters, M., Radelof, U. and Schneider, D.
I.M.A.G.E. cDNA Clone Collection
Unpublished (2004)
Contact: Inge Arlart
RZPD-Deutsches Ressourcenzentrum fuer Genomforschung GmbH
Heubnerweg 6, D-14059 Berlin, Germany
Email: www.rzpd.de
RZPD; IMAGE971N1484.
RZPDLIB; I.M.A.G.E. cDNA Clone Collection;
Contact: Inge Arlart
RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH
Heubnerweg 6, D-14059 Berlin, Germany
Tel: +49 30 32639 100
Fax: +49 30 32639 111
www.rzpd.de
This clone is available royalty-free from RZPD;
contact RZPD (clone@rzpd.de) for further information. Seq primer:
M13r, Primer sequence: TTTCACAGCAAGCAAGCTATGAC.

FEATURES
Location/Qualifiers
1..494
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE971N1484 ; IMAGE:2010852"
/tissue_type="gastroblastoma (pooled)"
/lab_host="DH10B"
/clone_lib="NCI CGAP Brn23"
/note="Organ: brain; Vector: pT73D-Pac (Pharmacia) with a
modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5',
TGTACCAATCTGAGTGGAGCGGCCCATATCTTTTCTTTTCTTTTCTTTT
T 3']; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
```

the Not I and Eco RI sites of the modified pT7T3 vector.
Library is normalized, and was constructed by Bento
Soares and M.Fatima Bonaldo."

```
ORIGIN
Query Match      38.3%; Score 493; DB 7; Length 494;
Best Local Similarity 99.8%; Pred. No. 6.5e-124;
Matches 493; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 299 AGAACAGCGATCAGCCGATTCAGCTGACCAAGTCCAAAGCAAGCAAGCGAGTCTCTCCG 358
Db 1 AGACAGCGATCAGCCGATTCAGCTGACCAAGTCCAAAGCAAGCAAGCGAGTCTCTCCG 60
QY 359 AAGCACAATCTTGTATGTCTCCACCTCAGAACGACGCTCTGTCTGATCGCCGACATGG 418
Db 61 AAGCACAATCTTGTATGTCTCCACCTCAGAACGACGCTCTGTCTGATCGCCGACATGG 120
QY 419 TCNAAGTCTCTCCCAAGGACACACCCCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 478
Db 121 TCNAAGTCTCTCCCAAGGACACACCCCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 180
QY 479 TGAAGCTGGAATGGAATGGAATGGAATGGAATGGAATGGAATGGAATGGAATGGAATG 538
Db 181 TGAAGCTGGAATGGAATGGAATGGAATGGAATGGAATGGAATGGAATGGAATGGAATG 240
QY 539 ATAAAGAGAAAGGCGGCGAGTCCAACTGGAATCCTCAGCATCTTCTGATTTACAGGCC 598
Db 241 ATAAAGAGAAAGGCGGCGAGTCCAACTGGAATCCTCAGCATCTTCTGATTTACAGGCC 300
QY 599 AGTTTCCTCGAGCCTCTCCAGACATCAGAGGCGAAATACCTGCTGTGATCTGGGCC 658
Db 301 AGTTTCCTCGAGCCTCTCCAGACATCAGAGGCGAAATACCTGCTGTGATCTGGGCC 360
QY 659 CACAAGAGGATGCAAAATCTCTAAGTTTACGGAGTCTCAATGACCATATCAGTCACT 718
Db 361 CACAAGAGGATGCAAAATCTCTAAGTTTACGGAGTCTCAATGACCATATCAGTCACT 420
QY 719 GGCTGCCCAACGTCAGAGTACCAAGTACCAAGTACCAAGTACCAAGTACCAAGTACCA 778
Db 421 GGCTGCCCAACGTCAGAGTACCAAGTACCAAGTACCAAGTACCAAGTACCAAGTACCA 480
QY 779 TGGACAAAGGCCAC 792
Db 481 TGGACAAAGGCCAC 494

RESULT 12
AA873869/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

AA873869 477 bp mRNA linear EST 17-MAR-1998
Ch79d06.s1 NCI_CGAP_Kid3 Homo sapiens cdna clone IMAGE:1473227 3',
mRNA sequence.
AA873869
AA873869.1 GI:2968005
EST.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 477)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
```

Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 471.
Location/Qualifiers

FEATURES
source

```
1. 477
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:1473227"
/lab_host="DH10B"
/clone_lib="NCI CGAP Kid3"
/note="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer,
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not
I and Eco RI sites of the modified pT7T3 vector. mRNA
source: 2 pooled kidneys. Library went through one round
of normalization. Library constructed by Bento Soares and
M. Fatima Bonaldo."
```

ORIGIN

```
Query Match      36.9%; Score 475.4; DB 1; Length 477;
Best Local Similarity 99.8%; Pred. No. 4.4e-119;
Matches 476; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 789 CCACCCCATCTTTTATTTGCAAGTCACTGTGCTCCAGATTCAGAACCCCTTCTACCTACAT 848
Db 477 CCACCCCATCTTTTATTTGCAAGTCACTGTGCTCCAGATTCAGAACCCCTTCTACCTACAT 418
QY 849 CAGTCACCTAGAAATCTCAGCTGGGTTTCAAAATGAAGACATGACCCGTTGTCTAGTGA 908
Db 417 CAGTCACCTAGAAATCTCAGCTGGGTTTCAAAATGAAGACATGACCCGTTGTCTAGTGA 358
QY 909 CCAGCAAGAGAGGAGGAGAGAGATCTCCCGGGTATCGTCGGCTCAGAGGTCTCCAGA 968
Db 357 CCAGCAAGAGAGGAGGAGAGAGATCTCCCGGGTATCGTCGGTTCAGAGGTCTCCAGA 298
QY 969 AACAAATAGTGTCCGGAAGAGGACACAGACTCTAAATTTCAAGTGTAAAGTTGTGTCTGG 1028
Db 297 AACAAATAGTGTCCGGAAGAGGACACAGACTCTAAATTTCAAGTGTAAAGTTGTGTCTGG 238
QY 1029 ATTTGTGAGCAAAACATGGGTAAACTCCACTAGCAAAAGCAGCAGCAAGTCAACCGA 1088
Db 237 ATTTGTGAGCAAAACATGGGTAAACTCCACTAGCAAAAGCAGCAGCAAGTCAACCGA 178
QY 1089 ACACCATTCACAGTTTGTAAACAGAGTGTGATCAAGAAATAGCTCTGAGGACGAATGCC 1148
Db 177 ACACCATTCACAGTTTGTAAACAGAGTGTGATCAAGAAATAGCTCTGAGGACGAATGCC 118
QY 1149 AGTTTCCACTTTCCAGCCTGGATCCCTCAGACTGAACCTTTCTTGGTTGCACCAATCTG 1208
Db 117 AGTTTCCACTTTCCAGCCTGGATCCCTCAGACTGAACCTTTCTTGGTTGCACCAATCTG 58
QY 1209 CTTCTGACATTTGAACCTCAATGAATCTCCTGACACCTTGGCTCTGAGAAGACTGCC 1265
Db 57 CTTCTGACATTTGAACCTCAATGAATCTCCTGACACCTTGGCTCTGAGAAGACTGCC 1
```

RESULT 13

```
CK773713/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS

CK773713 686 bp mRNA linear EST 20-FEB-2004
962388 MARC 2B0V Bos taurus cdna 3', mRNA sequence.
CK773713
CK773713.1 GI:42727856
EST.
Bos taurus (cow)
Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovinae; Bos.
1 (bases 1 to 686)
Smith,I.P.I., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C.,
```

Bennett, G.L., Heaton, M.P., Laegreid, W.W., Rohrer, G.A.,
Chitko-McKown, C.G., Pettea, G., Holt, I., Karamycheva, S., Liang, F.,
Quackenbush, J., and Keele, J.W.
Sequence evaluation of four pooled-tissue normalized bovine cdna
libraries and construction of a gene index for cattle
Genome Res. 11 (4), 626-630 (2001)

JOURNAL MEDLINE PUBMED

21180013
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smithemail.marc.usda.gov
Single pass sequencing. Bases called with phred v0.020425.c and
trimmed with the aid of the trim_alt option. Vector identified with
cross match v0.990329.

Plate: 76 row: K column: 24
Seq primer: GTAATACGATCACTATAGG.

FEATURES source

1. .686
Location/Qualifiers
/organism="Bos taurus"
/mol_type="mRNA"
/db_xref="taxon:9913"
/tissue_type="pooled"
/lab_host="DH10B"
/clone_lib="MARC 2BOV"
/note="Vector: pCMV SPORT6; Site 1: NotI; Site 2: SalI;
Library made from pooled tissue from testis, thymus,
semitendinosus muscle, longissimus muscle, pancreas,
adrenal, and endometrium."

ORIGIN

Query Match 36.1%; Score 465.2; DB 7; Length 686;
Best Local Similarity 88.2%; Pred. No. 3e-116;
Matches 506; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

Qy 444 CCCAAGCCAGCCTCTCTCCAGGGTCCCCCAGTGAAGCTGGAAATGGATGTGAGCGG 503
Db |||||
686 CCCGAAGCCCGCCACCTCCCGCAGGGTCCCTCCGTAAGCTGGAGATGGATGTGAGCGG 627
Qy 504 CTTGAGATGTCTCCAGTGAAGTCTCACTTTGCATTAAGAAAGAGCGGCGAGTCAAA 563
Db |||||
626 CTTGAGGACGTGTCCAGCGAGGTCTCCACCTTGATGAAGAGGAGCGGCGAGTCCAA 567
Qy 564 CTGGATCTCAGCATCTCTGATCTCAAGCCAGTTTGCTCGAGCCTCTCCAGAC 623
Db |||||
566 CTGGAAACCCCGAGCATCTCTGATCTCGAGGCTCAGTTTGCTCCAGCCTCTCCAGAC 507
Qy 624 ATCAGAGGGCAATACCTGCTCTGATCTGGGCCCCAAGAGCGGTATGCAAAATCTCTAA 683
Db |||||
506 CTCGAGGGCAATACCTGCTGTCCGACCTGGGCCCCGAGAGCGGATGCGATCTCGAA 447
Qy 684 GTTTACGGGACTCTCAATGACCACTATCAGTCACTGGCTGGCCAAAGCTCAAGTACCA 743
Db |||||
446 GTTTACGGGACTCTCGATGACCACTATCAGTCACTGGCTGGCCAAAGCTCAAGTACCA 387
Qy 744 TAGGAAACGGGCGGACAAATTTCTGAAACATGACAAAGCCACCCCATCTTTTA 803
Db |||||
386 TAGGAAACGGGCGGACCAAGTCTCTGAAACATGACAAAGCCACCCCATCTTTTA 327
Qy 804 TTGCAAGTGACTGTGCTCCAGTTTCAAGAACCCCTTTCTACCTACATCAGTCACTTGAATC 863
Db |||||
326 CTGCAAGTGACTGTGCTCCAGTTTCCGAACCCCGTCTACCTACATCAGCCACTTAGATC 267
Qy 864 TCACCTGGGTTTCCAAATGAAGGACATGACCCGCTTGTGTCAGTGGACGAGCAAGCAAGG 923
Db |||||
266 CCACCTGGGCTTCCAGATGAAGGACATGACCCGCTGTGTCAGTGGACGAGCAAGGCAAGC 207
Qy 924 GGAGCAAGGATCTCCCGGTTATCTCGGCTCAGAGGCTCCAGAAACATAGCTCCGGA 983
Db |||||
206 GGAGCAGGAGATCTCCCGGTTATCTCGGCTCAGCGGTCCCGGAAAGCAGCAGCTGGCGA 147

Qy 984 AGAGGACACAGACTCTAAATTCAGTGTAGTTG 1017
Db |||||
146 AGAGGACACAGGACTCTAAATTCAGTGTAGTTG 113

RESULT 14

LOCUS BO715061 1221 bp mRNA linear EST 16-JUL-2002
DEFINITION AGENCOURT_8474842 NIH_MGC_129 Mus musculus cdna clone IMAGE:6310221
5', mRNA sequence.

ACCESSION BO715061

VERSION BO715061.1 GI:21853960

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM

Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 1221)

NIH-MGC <http://mgi.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

Tissue Procurement: Susan L. Sullivan, Ph.D.

cDNA Library Preparation: Resgen, Invitrogen Corp

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: LLAM13732 row: c column: 22

High quality sequence start: 79

High quality sequence stop: 516.

FEATURES

source

1. 1221

Location/Qualifiers

/organism="Mus musculus"

/mol_type="mRNA"

/db_xref="taxon:10090"

/clones="IMAGE:6310221"

/lab_host="DH10B (phage-resistant)"

/clone_lib="NIH MGC 129"

/note="Organ: olfactory epithelium; Vector: pCMV-SPORT6.1;

Site 1: EcoRV; Site 2: NotI; Cloned unidirectionally.

Primer: Oligo dt. Average insert size 2.2 kb. Constructed

by Resgen, Invitrogen Corp. Note: this is a NIH_MGC

ORIGIN

Query Match 34.0%; Score 438; DB 5; Length 1221;
Best Local Similarity 88.0%; Pred. No. 1e-108;
Matches 500; Conservative 0; Mismatches 65; Indels 3; Gaps 2;

Qy 331 TCCAAAGCAGAGAGCGAGTCTCGAAGCACAATCTTGTATGTCCTCCCACTCAGAG 390
Db |||||
91 TCCAAAGCAGAGAGCGAGTCTCGAAGCACAATCTTGTATGTCCTCCCACTCAGAG 150
Qy 391 CACGCTCTGTGTGACATCGCCGACATGTCTCAAGTCTCTCCCAAGCCACCACTCAGAG 450
Db |||||
151 CATGCTCTGTGTGACATGTGTGATGTCTCAAGTCTCTCCCAAGCCACCACTCAGAG 210
Qy 451 CCAGCTCTCTCTCAGGGTCCCCCATGAAGCTGGAAATGGATGTGAGGGCTTTGAG 510
Db |||||
211 CCAGCGGCTCTCAGGGTCCCTCTATGAAGCTGGAAATCGACGTGAGCGTTCCGAG 270
Qy 511 GATGCTCCAGTGAAGTCTCACTTTGCATTAAGAAAGCGGCGAGTCCCACTGGAGT 570
Db |||||
271 GATGTTCCAGGAGTCTCCACTTTGCACAAAAGAAAGCGGCGAGTCCCACTGGAGT 330
Qy 571 CCTCAGCATCTTCTGATTTCAAGCCAGTTTGTCTCGAGCCTCTTCCAGACATCAGAG 630
Db |||||
331 CCCAGCATCTTCTCATCTCGAAGCTCAGTTTGTCTCGAGCCTCTTCCAGACCTCAGAG 390
Qy 631 GGCAATACCTGCTGTCTGATCTGGGCCCCAAGAGCGTATGCAAAATCTTAAAGTTTACG 690

Db	391	GGCAAAATACCTCTTCTGACCTGGGCCCCCAAGAGCGGATGCAAAATCTCAAGTTCCACC	450
Qy	691	GGACTCTCAATGACCACTATCAGTCACTGGCTGGCCAAACGTCAGTAGTACCAGCTTAGGAAA	750
Db	451	GAATCTCCTCATGACCAATCATGACCTGCTGGCTGGCTAAACGTCAGTAATACCAAGCTTAGGAAA	510
Qy	751	ACGGGGGGGACAAAATTTCTGAAAACATGACAAAGGCCACCCCATCTTTTATGTCAGT	810
Db	511	ACAGGTGGGACAAAGTTCTGTAAGAAATGACAAAGGGCACCCCATCTTTTACTGTCAGC	570
Qy	811	GACTGTGCTCCAGTTTCAGAA-CCCTTCTACCTACATAGTCACTTAGATCTCACT	869
Db	571	GACTGTGCTCCAGTTTCAGAA-CCCTTCTACCTACATAGTCACTTAGATCTCACT	630
Qy	870	GGG--TTTCCAAATGAAGACATGACCC	895
Db	631	GGGGCTTCCCAATGAAGACATGACCC	658
LOCUS	CK775937	581 bp	linear
DEFINITION	967078 MARC 380V Bos taurus cDNA 3', mRNA	EST 20-FEB-2004	
ACCESSION	CK775937		
VERSION	CK775937.1	GI:42732187	
KEYWORDS	EST.		
SOURCE	Bos taurus (cow)		
ORGANISM	Bos taurus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.		
AUTHORS	Smith, T.P.L., Grosse, W.M., Freking, B.A., Roberts, A.J., Stone, R.T., Casas, E., Wray, J.E., White, J., Cho, J., Fahrenkrug, S.C., Bennett, G.L., Heaton, M.P., Laegreid, W.W., Rohrer, G.A., Chitko-McKown, C.G., Perlea, G., Holt, I., Karamycheva, S., Liang, F., Quackenbush, J. and Keefe, J.W.		
TITLE	Sequence evaluation of four pooled-tissue normalized bovine cDNA libraries and construction of a gene index for cattle		
JOURNAL	Genome Res. 11 (4), 626-630 (2001)		
MEDLINE	21180013		
PUBMED	11282978		
COMMENT	Contact: Smith TPL USDA, ARS, US Meat Animal Research Center PO Box 166, Clay Center, NE 68933-0166, USA Tel: 402 762 4366 Fax: 402 762 4390 Email: smith@email.marc.usda.gov Single pass sequencing. Bases called with phred v0.020425.c and trimmed with the aid of the trim_alt option. Vector identified with cross_match v0.990329. Plate: 119 row: P column: 16 Seq primer: GTAATACGACTCACTATAGG.		
FEATURES	Location/Qualifiers		
source	1..581 /organism="Bos taurus" /mol_type="mRNA" /db_xref="taxon:9913" /tissue_type="pooled" /lab_host="DH108" /clone_lib="MARC 380V" /note="Vector: pCMV SPORT6; Site 1: NotI; Site 2: SalI; Library made from pooled tissue from marrow, alveolar macrophage, ovary, fetal semitendinosus muscle, and fetal longissimus muscle."		
ORIGIN			
Query Match	33.8%;	Score 435.2;	DB 7; Length 581;
Best Local Similarity	86.1%;	Pred. No. 5e-108;	
Matches 482;	Conservative 0;	Mismatches 78;	Indels 0; Gaps 0;
Qy	676	ATCTCTAAGTTTACGGGACTCTCAATGACCACTATGCTGCTGGCCAAAGCTCAAG	735

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 4, 2005, 20:11:25 ; Search time 2220.75 Seconds
(without alignments)
9927.764 Million cell updates/sec

Title: US-08-731-499-8
Perfect score: 455
Sequence: 1 GAAATCAGAGCTTTAATATG.....ATTGCTCTATTTCNGGGGGT 455

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb.ba.*
- 2: gb.htg.*
- 3: gb.in.*
- 4: gb.on.*
- 5: gb.ov.*
- 6: gb.pat.*
- 7: gb.ph.*
- 8: gb.pl.*
- 9: gb.pr.*
- 10: gb.ro.*
- 11: gb.ars.*
- 12: gb.sy.*
- 13: gb.un.*
- 14: gb.vl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	453	99.6	455	6 AR070331	Sequence
2	453	99.6	455	6 BD085732	Genes fro
3	428	94.1	4723	9 BC071570	Homo sapi
4	428	94.1	4944	9 AB002385	Human mRN
5	428	94.1	118947	2 AC006321	Homo sapi
6	426.4	93.7	2328	6 AR369583	Sequence
7	426.4	93.7	2328	6 AR404162	Sequence
8	426	93.6	4699	9 HS081561	Human prote
9	424.8	93.4	190846	9 AC006372	Homo sapi
10	416	91.4	4735	9 BC034040	Homo sapi
11	414.6	91.1	4719	9 HS06702	Human phogr
12	105	23.1	105	11 GI9707	human STS A
13	86.2	18.9	180187	2 AC117965	Rattus no
14	85.6	18.8	5136	10 RNU73458	Rattus norv
15	76.8	16.9	222251	2 AC110936	Rattus no
16	74.6	16.4	197934	2 AC117676	Mus muscu
17	74.6	16.4	211959	2 AC118023	Mus muscu
18	50.4	11.1	105604	9 AC084013	Homo sapi
19	49.8	10.9	106216	9 AC061997	Homo sapi

20	49.8	10.9	255522	2	AC129169	Rattus no
21	49.4	10.9	51	6	CQ003853	Sequence
c 22	49	10.8	249943	3	AE014823	Plasmodi
c 23	48	10.5	75076	9	AC004948	Homo sapi
c 24	47.8	10.5	103344	9	HS1100E15	Human DNA
c 25	46.6	10.2	82139	3	AC115684	Dicystoate
c 26	46.6	10.2	149591	2	AC141821	Apis mell
c 27	46.4	10.2	191481	2	AC108123	Homo sapi
c 28	46	10.1	144759	9	HS352A20	Human DNA
c 29	46	10.1	162905	2	EX927388	Danio rer
c 30	45.8	10.1	127230	9	AC027296	Homo sapi
c 31	45.8	10.1	173997	2	AC012340	Homo sapi
c 32	45.8	10.1	176898	9	AC117569	Homo sapi
c 33	45.8	10.1	194028	2	AC111077	Homo sapi
c 34	45.8	10.1	229569	2	AC107424	Homo sapi
c 35	45.6	10.0	158548	3	PFM43P2	Plasmodi
c 36	45.6	10.0	272384	2	AC125722	Rattus no
c 37	45.4	10.0	2000	6	AX655393	Sequence
c 38	45.4	10.0	150754	9	AC023491	Homo sapi
c 39	45.4	10.0	152884	9	AC138625	Homo sapi
c 40	45.4	10.0	156168	2	EX927212	Danio rer
c 41	45.4	10.0	158420	9	AC137788	Homo sapi
c 42	45.4	10.0	175358	9	AC007981	Homo sapi
c 43	45.4	10.0	192049	2	AC137785	Homo sapi
c 44	45.2	9.9	770	11	BV006800	STS_CBB26
c 45	45.2	9.9	8099	9	AP001114	Homo sapi

ALIGNMENTS

RESULT 1
AR070331
LOCUS AR070331 455 bp DNA linear PAT 18-FEB-2000
DEFINITION Sequence 8 from patent US 5892010.
ACCESSION AR070331
VERSION AR070331.1 GI:7221219
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 455)
AUTHORS Gray,J., Collins,C., Hwang,S.-i., Godfrey,T., Kowbel,D. and Rommens,J.
TITLE Genes from the 20Q13 amplicon and their uses
JOURNAL Patent: US 5892010-A 8 06-APR-1999;
FEATURES Location/Qualifiers
source 1..455
/organism="unknown"
/mol_type="unassigned DNA"

ORIGIN	Query Match	99.6%	Score 453;	DB 6;	Length 455;
	Best Local Similarity	100.0%	Pred. No. 6.2e-102;	Mismatches 0;	Indels 0; Gaps 0;
	Matches	455;	Conservative 0;		
Qy	1	GAAATCAGAGCTTTAATATGACACAAATTAATATATTTTGTATATCTCATCCGGAGNTTC	60		
Db	1	GAAATCAGAGCTTTAATATGACACAAATTAATATATTTTGTATATCTCATCCGGAGNTTC	60		
Qy	61	TCCTTCAACATAGAGGTTAGAAATTAACAAGTAGGCATATGCTTCCTATATTCAGATAAA	120		
Db	61	TCCTTCAACATAGAGGTTAGAAATTAACAAGTAGGCATATGCTTCCTATATTCAGATAAA	120		
Qy	121	TTTCATTCGATTAAATTAATTCAGATAGAGAGTAATTTTCGAAAAGAAATGATAG	180		
Db	121	TTTCATTCGATTAAATTAATTCAGATAGAGAGTAATTTTCGAAAAGAAATGATAG	180		
Qy	181	CTATATTAAAGCAGATATTCATTACAATACCATGTAGAGACATAGCAATATTTTGGCAT	240		
Db	181	CTATATTAAAGCAGATATTCATTACAATACCATGTAGAGACATAGCAATATTTTGGCAT	240		
Qy	241	CATTCTGTCCGCTCAGTAGGGCGGTTCCTCTCTGGTGGGCGCTTTGGAGAGTACCATCTA	300		

```

Db      241  |||||TGTCCGCTCAGTAGGCCGCTGTTCCCTCTGGTAGGCCCTTTGGAGAGTACCATCTA 300
Qy      301  TCTAAGATGGAGAAATGCTGTGGAGAGGGCGGATGGAGGTGGTTCCTACGCTGAACC 360
Db      301  TCTAAGATGGAGAAATGCTGTGGAGAGGGCGGATGGAGGTGGTTCCTACGCTGAACC 360
Qy      361  CCACACAGGAAATCTGCAGGCCACACAGCTGCTCTGCGCGCGCTTCCATGTGATCATCC 420
Db      361  CCACACAGGAAATCTGCAGGCCACACAGCTGCTCTGCGCGCGCTTCCATGTGATCATCC 420
Qy      421  TGGTCAATGAAGTGAATTCCTATTTCNGGGGGT 455
Db      421  TGGTCAATGAAGTGAATTCCTATTTCNGGGGGT 455

RESULT 2
BD085732
LOCUS   Genes from the 20q13 amplicon and their uses. PAT 27-AUG-2002
DEFINITION
ACCESSION BD085732
VERSION   JP 2001524802-A/8
KEYWORDS  synthetic construct
SOURCE    synthetic construct
ORGANISM  other sequences; artificial sequences.
REFERENCE 1 (bases 1 to 455)
AUTHORS   Gray,J.W., Collins,C.C., Hwang,S.I., Godfrey,T., Kowbel,D. and Rommens,J.
TITLE     Genes from the 20q13 amplicon and their uses
JOURNAL   Patent: JP 2001524802-A 8 04-DEC-2001;
COMMENT   THE REGENTS OF THE UNIVERSITY OF CALIFORNIA
OS        Artificial Sequence
PN        JP 2001524802-A/8
PD        04-DEC-2001
PF        15-JUL-1997 JP 1998506264
PR        15-JUL-1996 US 08/680395,16-OCT-1996 US 08/731499 PR
PI        17-JAN-1997 US 08/785532
PI        JOE W GRAY, COLIN CONRAD COLLINS, SOO IN HWANG, TONY GODFREY, PI
DAVID KOWBEL,
PI        JOHANNA ROMMENS
PC        C12N15/11.C12Q1/68.A61K48/00
CC        Description of Artificial Sequence:20sa7
CC        n is A, C, G, T, or U
FH        Key Location/Qualifiers
FT        modified base (57)
FT        modified base (449).
FEATURES
source    1..455
           /organism="synthetic construct"
           /mol_type="genomic DNA"
           /db_xref="taxon:32630"

ORIGIN
Query Match 99.6%; Score 453; DB 6; Length 455;
Best Local Similarity 100.0%; Pred. No. 6.2e-102;
Matches 455; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1  GAAATCAGAAGTTTAATATGACACAAATTAATATTTGTATATCTCACACCGAGNTTC 60
Db      1  GAAATCAGAAGTTTAATATGACACAAATTAATATTTGTATATCTCACACCGAGNTTC 60

Qy      61  TCTTCAACATAGGAGTTAGAAATACAGTAGGCATATGCTTCCTATATTCAGATAAA 120
Db      61  TCTTCAACATAGGAGTTAGAAATACAGTAGGCATATGCTTCCTATATTCAGATAAA 120

Qy      121  TTCATTTCGATTAAATTAATTCAGATAGAGAGAAGTAATTTTCGGAAAGAAATCATAG 180
Db      121  TTCATTTCGATTAAATTAATTCAGATAGAGAGAAGTAATTTTCGGAAAGAAATCATAG 180

Qy      181  CTATATTAAAGCAGATATTCATTACAAATACCATGTAGACATACAAATATTTTGGCAT 240
Db      181  CTATATTAAAGCAGATATTCATTACAAATACCATGTAGACATACAAATATTTTGGCAT 240
```

```

Qy      241  CATTTCTGTCGCTCAGTAGGCCGCTGTTCCCTCTGGTAGGCCCTTTGGAGAGTACCATCTA 300
Db      241  CATTTCTGTCGCTCAGTAGGCCGCTGTTCCCTCTGGTAGGCCCTTTGGAGAGTACCATCTA 300
Qy      301  TCTAAGATGGAGAAATGCTGTGGAGAGGGCGGATGGAGGTGGTTCCTACGCTGAACC 360
Db      301  TCTAAGATGGAGAAATGCTGTGGAGAGGGCGGATGGAGGTGGTTCCTACGCTGAACC 360
Qy      361  CCACACAGGAAATCTGCAGGCCACACAGCTGCTCTGCGCGCGCTTCCATGTGATCATCC 420
Db      361  CCACACAGGAAATCTGCAGGCCACACAGCTGCTCTGCGCGCGCTTCCATGTGATCATCC 420
Qy      421  TGGTCAATGAAGTGAATTCCTATTTCNGGGGGT 455
Db      421  TGGTCAATGAAGTGAATTCCTATTTCNGGGGGT 455

RESULT 3
BC071570/c
LOCUS   4723 bp mRNA linear PRI 25-JUN-2004
DEFINITION Homo sapiens protein tyrosine phosphatase, receptor type, N
polyptide 2, mRNA (cdna clone MGC:87154 IMAGE:4375815), complete
cds.
ACCESSION BC071570
VERSION   BC071570.1 GI:47939488
KEYWORDS  MGC.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
REFERENCE 1 (bases 1 to 4723)
AUTHORS   Strausberg R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
Schetz,T.E., Brownstein,M.J., Ustin,T.B., Toshiyuki,S.,
Carninci,P., Frange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
Fahy,J., Helton,E., Kettman,M., Madan,A., Rodriguez,S.,
Sanchez,A., Whiting,M., Madan,A., Young,A.C., Greenchenko,Y.,
Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Shreenk,R.,
Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.B.,
Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
12477932
2 (bases 1 to 4723)
Strausberg,R.
Direct Submission
Submitted (01-JUN-2004) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgabs-r@mail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
cdNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki
Toshiyuki and Piero Carninci (RIKEN)
cdNA Library Arrayed by: The I.M.A.G.E. Consortium (LINL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome
Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: http://www.shgc.stanford.edu
Contact: (Dickson, Mark) mcd@paxil.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
```

R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAK Plate: 167 Row: b Column: 15
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 19743910.

FEATURES

source

1. 4723
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MGC:87154 IMAGE:4375815"
/tissue_type="Brain, hippocampus"
/clone_lib="NIH MGC_95"
/lab_host="DH10B"
/note="Vector: pBluescript"

gene

1. 4723
/gene="PTPRN2"
/note="synonyms: IA-2beta, IAR, ICAR, KIAA0387, PTPRP, IAR PTPRP"
/db_xref="LocusID:5799"
/db_xref="MIM:601698"

CDS

120..2021
/gene="PTPRN2"
/codon_start=1
/product="PTPRN2 protein"
/protein_id="AAH71570.1"
/db_xref="GI:47939489"
/db_xref="LocusID:5799"
/db_xref="MIM:601698"
/translation="MGPLPLLLLLPRVLPAAPSSVPRGRQLPGRIGCLLEEG
LCGASEACVNGFTWDDYTQVMDQELADLPKTYLRPEASSPARPKSHVSGSRRYS
REGGAALANLRULPFTLEYSQASDPAVLRTAQRPPAEGDPRFSEILTYVAH
TSALTYPPQPTOLREDLLPTLQLOPDELSPKVDGVDHRLHMAALSAVAARPPA
PPGSGLEPQVLLRAPSMPRLPAAPQKWPSPGSDPSSTGSGARHTLTKOL
QRQAENVRLSGLEDGMAGLMQGVHGVARGSPRAALGESGEGQADGPKATLR
GDSPPDDGVDDRLYQEVHRLSATLGLLQDHGSRLLPGALTFARPLDNERKKESEH
PESSLSSEETAGDENVKTSQYKDLGQPHSPGAAGFELQNMQPGPKSQSLP
AGAEALSDGLQLEVPQSEERARGYIVTDRLPAPERRRIVEDVARLLQVPSAFAD
VEVLGPATVFKVSANQVNTVEDVEKATVDNKKLEETSGLKILQTVGSGSKLFLP
FOAQEDSTKFIATLTVSLACILGVLLASGLIYCLRHSSQHRLEKSLGLGDPGADA
TAAYQELCR"

ORIGIN

Query Match 94.1%; Score 428; DB 9; Length 4723;
Best Local Similarity 99.1%; Pred. No. 1.2e-95;
Matches 450; Conservative 0; Mismatches 2; Indels 2; Gaps 2;

QY 1 GAAATCAGAAGTTTAATATGACACAAATTAATATATTTGTATATCTCACACCGGA-GNNT 59
DB 4705 GAAATCAGAAGTTTAATATGACACAAATTAATATATTTGTATATCTCACACCGGAGTTT 4646

QY 60 CTCCTTCAACATAAGAGGTAGAAATTAACAGTAGGCATATGCTTCCTATATTCAGATAA 119
DB 4645 CTCCTTCAACATAAGAGGTAGAAATTAACAGTAGGCATATGCTTCCTATATTCAGATAA 4586

QY 120 ATTCATTTTCGATTAAATTCAGATAGAGAGAGTAATTTTCGGAAGAAAGTAATGATA 179
DB 4585 ATTCATTTTCGATTAAATTCAGATAGAGAGAGTAATTTTCGGAAGAAAGTAATGATA 4527

QY 180 GCTATATTAAAGCAGATATTCATTACATACCATGTAGACATAAGCAATATTTTGGCA 239
DB 4526 GCTATATTAAAGCAGATATTCATTACATACCATGTAGACATAAGCAATATTTTGGCA 4467

QY 240 TCATTCTGTCGCTCAGTAGGCGGTGTTCCCTCTCGTAGGCGCTTTGGAGAGTACCATCT 299
DB 4466 TCATTCTGTCGCTCAGTAGGCGGTGTTCCCTCTCGTAGGCGCTTTGGAGAGTACCATCT 4407

QY 300 ATCTAAGATGGAGGAATGCTGTGGGAAGCGCGGATGAGAGTGCGCTTTTCTACGCTGAAC 359
DB 4406 ATCTAAGATGGAGGAATGCTGTGGGAAGCGCGGATGAGAGTGCGCTTTTCTACGCTGAAC 4347

ORIGIN	SDCAGRSCTVLTLDMLNVKAKAKEIDIAATLEHLRDQRPGMVQTKQEPFAATAVA EEVNALIKALPQ"	
Query Match	94.1%	Score 428; DB 9; Length 4944;
Best Local Similarity	99.1%	Pred. No. 1.2e-95;
Matches 450; Conservative	0; Mismatches 2; Indels 2; Gaps 2;	
Qy	1	GAATCAGAAAGTTTAATATGACACAAATTAATAATATATTGTATATCTCACACCGGA-GNTT 59
Db	4942	GAATCAGAAAGTTTAATATGACACAAATTAATAATATATTGTATATCTCACACCGGAGTTT 4883
Qy	60	CTCTTCAAAATGAAGGAGTTAGAAATPACAAGTAGGCATATGCTTCTATATTCAGATAA 119
Db	4882	CTCTTCAAAATGAAGGAGTTAGAAATPACAAGTAGGCATATGCTTCTATATTCAGATAA 4823
Qy	120	ATTCAATTTCCATTAATTTAAATTCAGATAGAGAGAGTAATTTTCGAAAAGAAATGATA 179
Db	4822	ATTCAATTTCCATTAATTTAAATTCAGATAGAGAGAGTAATTTT-GGAAAAGAAATGATA 4764
Qy	180	GCTATATTAAGCAGATATTCATTACAAATACCATGTAGACACATACAAATATTTTGGCA 239
Db	4763	GCTATATTAAGCAGATATTCATTACAAATACCATGTAGACACATACAAATATTTTGGCA 4704
Qy	240	TCATTCGTCGCTCAGTAGCCGCTGTTCCCTCTGTGTAAGGCCCTTTGGAGAGTACCATCT 299
Db	4703	TCATTCGTCGCTCAGTAGCCGCTGTTCCCTCTGTGTAAGGCCCTTTGGAGAGTACCATCT 4644
Qy	300	ATCTAGATCGAGGATGCTGTGGGAGGGGGATGGAGTGCGTTTCTACGCTGAAC 359
Db	4643	ATCTAGATCGAGGATGCTGTGGGAGGGGGATGGAGTGCGTTTCTACGCTGAAC 4584
Qy	360	CCACACAGGAAATCTCAGCCACACAGCTGCTCTGCGCGCCTTCCATGTGATCATC 419
Db	4583	CCACACAGGAAATCTCAGCCACACAGCTGCTCTGCGCGCCTTCCATGTGATCATC 4524
Qy	420	CTGGTCAATGAAGTGAATTCCTATTTCNGGG 453
Db	4523	CTGGTCAATGAAGTGAATTCCTATTTCNGGG 4490
RESULT 5		
AC006321	AC006321 118947 bp DNA linear HTG 24-AUG-2000	
LOCUS	Homo sapiens chromosome UNK clone RP5-991P5, WORKING DRAFT	
DEFINITION	SEQUENCE, 2 uncloned pieces.	
ACCESSION	AC006321	
VERSION	AC006321.3 GI:9887769	
KEYWORDS	HTG; HTGS, PHASE1; HTGS_DRAFT.	
SOURCE	Homo sapiens (human)	
ORGANISM	Homo sapiens	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
JOURNAL	Waterston, R.H.	
REFERENCE	The sequence of Homo sapiens clone	
AUTHORS	Waterston, R.H.	
JOURNAL	Unpublished	
REFERENCE	2 (bases 1 to 118947)	
AUTHORS	Waterston, R.H.	
JOURNAL	Direct Submission	
REFERENCE	Submitted (09-JAN-1999) Genome Sequencing Center, Washington	
AUTHORS	University School of Medicine, 4444 Forest Park Parkway, St. Louis,	
JOURNAL	MO 63108, USA	
COMMENT	On Aug 24, 2000 this sequence version replaced gi:8018216.	
	----- Genome Center -----	
	Center: Washington University Genome Sequencing Center	
	Center code: WUGSC	
	Web site: http://genome.wustl.edu/gsc/index.shtml	
	----- Project Information -----	
	Center project name: H DU099IP05	
	----- Summary Statistics -----	
	Sequencing vector: M13; 85%	

Sequencing vector: plasmid; 15%	
Chemistry: Dye-terminator ET; 85% of reads	
Chemistry: Dye-terminator Big Dye; 15% of reads	
Assembly program: Phrap; version 0.990319	
Consensus quality: 117806 bases at least Q40	
Consensus quality: 118306 bases at least Q30	
Consensus quality: 118463 bases at least Q20	
Insert size: 132000; agarose-fp	
Insert size: 118847; sum-of-contigs	
Quality coverage: in Q20 bases; sum-of-contigs	

* NOTE: This is a 'working draft' sequence. It currently	
* consists of 2 contigs. The true order of the pieces	
* is not known and their order in this sequence record is	
* arbitrary. Gaps between the contigs are represented as	
* runs of N, but the exact sizes of the gaps are unknown.	
* This record will be updated with the finished sequence	
* as soon as it is available and the accession number will	
* be preserved.	

* 1 92854: contig of 92854 bp in length	
* 92855 92954: gap of unknown length	
* 92955 118947: contig of 25993 bp in length.	

FEATURES	
source	1. 118947
	/organism="Homo sapiens"
	/mol_type="genomic DNA"
	/db_xref="taxon:9606"
	/chromosome="UNK"
	/clones="RP5-991P5"
misc_feature	1. 92854
	/note="assembly_name:Contig10
	clone end:SP6
	vector_side:right"
misc_feature	92955..118947
	/note="assembly_name:Contig9
	clone end:T7
	vector_side:left"

ORIGIN	
Query Match	94.1%; Score 428; DB 2; Length 118947;
Best Local Similarity	99.1%; Pred. No.1.4e-95;
Matches 450; Conservative 0; Mismatches 2; Indels 2; G	
QY	1 GAAATCAGAAGTTAAATATCACACAATTAAATATATTTGTTATATCTCACACGGGAA
Db	29867 GAAATCAGAAGTTAAATATCACACAATTAAATATATTTGTTATATCTCACACGGGAG
QY	60 CTCTTCAACATAGGAGTTAGAAATTTACAAGTAGGCATATGCTTCTTATATTCAGAG
Db	29927 CTCTTCAACATAGGAGTTAGAAATTTACAAGTAGGCATATGCTTCTTATATTCAG
QY	120 ATTCAATTCGATTAATTTAAATTCAGATAGAGAAAGTAATTTTCGGAAAGAAAT
Db	29987 ATTCAATTCGATTAATTTAAATTCAGATAGAGAAAGTAATTTT-GGAAAAAGAAAT
QY	180 GCTATATTAAGCAGATATTCATTCAATACCATGTAGAGACATAAGCAATATTTT
Db	30046 GCTATATTAAGCAGATATTCATTCAATACCATGTAGAGACATAAGCAATATTTT
QY	240 TCATTCTGTCGCTCAGTAGGCGGTGTTCCTCTGTTAGGCGCTTTGGAGAGTAGCC
Db	30106 TCATTCTGTCGCTCAGTAGGCGGTGTTCCTCTGTTAGGCGCTTTGGAGAGTAGCC
QY	300 ATCTAAGATGAGGAATGCTGTGGGAAGGCGGGAATGGAGGTGCGTTTTTCTACGCT
Db	30166 ATCTAAGATGAGGAATGCTGTGGGAAGGCGGGAATGGAGGTGCGTTTTTCTACGCT
QY	360 CCCACACAGGAAATCTCGACGCCACACAGCTGCCTCTGCGCGCCCTTCCATGTGAT
Db	30226 CCCACACAGGAAATCTCGACGCCACACAGCTGCCTCTGCGCGCCCTTCCATGTGAT
QY	420 CTGTGCAATGAAGTGAATTCCTCTATTTTCNCGGG 453

source	1. .4699 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /chromosomes="7" /map="7q22-qter" /tissue_type="brain; pancreas" 1. .4699 /gene="PTPRP" 43. .3039 /gene="PTPRP" /codon_start=1 /product="protein tyrosine phosphatase receptor pi" /protein_id="AAB68603.1" /db_xref="GI:2351576" /translation="MGPPPLPLLLLLLLLPPRVLPAAPSSVPRQLPGRLLDGVGRRC QKVPAMDFRYEYVSRQLRVAKLQSGTFTWDDYTYQYMQEALDLPKTYLR PRASSPFRYSKVSERYTREGGAALANLRHLFLFLEDSQAPASVLPARTHTAQ DRPPABDDRFESILTYVAHTSALTYPGPRTQLRDLPLRLTGLQLOPDELSPKVS GVDHHLMAALGAYAAQRPAPPPGEGSLPEVYLLRAPSRMPRLAPAPOKWPSPLG DSDPSSTGCGARIHTLLDLQRPAPVGLNGLELDGMAELMAGLMQGVDRGVRGS PGRALGESGEADGPKATLRGDSFPDGVQDDDKLYQEVHRLSATLGLGLQDGRSR LPLGALFPARLDMEKKKSEHPSSLSSEETAGVENKVSQYKDLGQQHSPFGA AAGFELQNPQPSKEEQSLPAGAQALSDGLQLEVPSEEARGYIVTDRPLRPEE GRRLVEDARLLQVPSAEADVEVLGPAVTFKVSANQVNTTEDVEKATVDMKNKLEE TSGLKLTQVGSKSLKFLPPQAEQDSTKFIATLVSLACTILGVLLASGLIYCLRH SSQHLRKLKSLGSDPGADATAAVQELCRQMATPDRPDRPGPHTSRSSVSQFSD GTPSPASRSSASSWEPEPQSNMDLSTGHMILSYMEDHLKNRLEKWEALCATQA ENSSFVAQRENVKPNRSLAVITYDHSRLVLAENSHSHSYDYNASPMIDHPRNPA YIATQGLPATVPMQWSEGVVIMLTPLAENGVRQCYHWPDEGSLNHYHYEV NLVSEHICWEDFLVRSFYLNKTNETRTVTFQFLSWYDRGVSPSSRLDPRKVN KCVRGSCPIIVHCSDGAGRSQTYVLIDVLNKMKAKEIDIAATLEHLRDRQPGW QTKQEPEFALTAABEVNAILKALPQ" 1834. .1908 /misc_feature /gene="PTPRP" /notes="encodes transmembrane domain" 2185. .3006 /misc_feature /gene="PTPRP" /note="encodes catalytic domain"
misc_feature	
misc_feature	
ORIGIN	
Query Match	93.6%; Score 426; DB 9; Length 4699;
Best Local Similarity	99.1%; Pred. No. 3.7e-95;
Matches	448; Conservative 0; Mismatches 2; Indels 2; Gaps 2;
QY	3 AATCAGAAGTTTAATATGACACAAATTAATATATTTGTATATCTCACCGGA-GNWTCT 61
Db	4690 AATCAGAAGTTTAATATGACACAAATTAATATATTTGTATATCTCACCGGAGTTTCT 4631
QY	62 CTTCAACATAGGAGTTAGAAATTAACAGTAGGCATATGCTTCTATATTCAGATAAAT 121
Db	4630 CTTCAACATAGGAGTTAGAAATTAACAGTAGGCATATGCTTCTATATTCAGATAAT 4571
QY	122 TCATTTCCGATTAATTAATTCAGATAGAGAGTAATTTTCGAAAAAGAAATGATAGC 181
Db	4570 TCATTTCCGATTAATTAATTCAGATAGAGAGTAATTTT-GGAAAAAGAAATGATAGC 4512
QY	182 TATATTAAAGCAGATATTCATTAACATACCATGATAGACATAGCAATATTTTGGCATC 241
Db	4511 TATATTAAAGCAGATATTCATTAACATACCATGATAGACATAGCAATATTTTGGCATC 4452
QY	242 ATTCTCTCGCTCAGTAGGCGGTTCCTCTGTTAGGCGCTTTGGAGAGTACCATCTAT 301
Db	4451 ATTCTCTCGCTCAGTAGGCGGTTCCTCTGTTAGGCGCTTTGGAGAGTACCATCTAT 4392
QY	302 CTAAGATGAGGAATGCTGTGGGAAGCGGGATGAGAGTGCGTTTCTACGCTGAACCC 361
Db	4391 CTAAGATGAGGAATGCTGTGGGAAGCGGGATGAGAGTGCGTTTCTACGCTGAACCC 4332
QY	362 CACACAGGAATTCGAGCCCAACAGCTGCCTCTCGCGCGCTTCCATGTGATCATCT 421
Db	4331 CACACAGGAATTCGAGCCCAACAGCTGCCTCTCGCGCGCTTCCATGTGATCATCT 4272
QY	422 GGTCAATGAAGTGAATTTGCTTATTTCTTTCGGGG 453
Db	4271 GGTCAATGAAGTGAATTTGCTTATTTCTTTCGGGG 4240
RESULT 9	
AC006372	
LOCUS	190846 bp DNA linear PRI 31-JAN-2004
DEFINITION	Homo sapiens BAC clone RP11-331D5 from 7, complete sequence.
ACCESSION	AC006372
VERSION	AC006372.2 GI:4753237
KEYWORDS	HTG.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 190846)
AUTHORS	Hillier,L.W., Fulton,R.S., Fulton,L.A., Graves,T.A., Pepin,K.H., Wagner-McPherson,C., Layman,D., Maas,J., Jaeger,S., Walker,R., Wylie,K., Sekhon,M., Becker,M.C., O'Laughlin,M.D., Schaller,M.E., Fewell,G.A., Delehaanty,K.D., Miner,T.L., Nash,W.E., Cordes,M., Isak,A., Vanbrunt,A., Nguyen,C., Du,F., Lamar,B., Ali,J., Andrews,S., Kalkbrenner,J., Edwards,J., Bradshaw-Cordum,H., Li,J., Courtney,L., Harkins,R., Harris,A., Strong,C.M., Hou,S., Tomlinson,C., Daughin-Kohlberg,S., Kozlowski-Reilly,A., Leonard,S., Rohlfing,T., Rock,S.M., Tin-Wollam,A.M., Abbott,A., Minx,P., Maupin,R., Stromwater,C., Latreille,P., Miller,N., Johnson,P., Murray,J., Woessner,J.P., Wendt,M.C., Yang,S.P., Schultz,B.R., Wallis,J.W., Speth,J., Bieri,T.A., Nelson,J.O., Berkowicz,N., Wohlmann,P.E., Cook,L.L., Hickenbotham,M.T., Eldred,J., Williams,D., Bedell,J.A., Mardis,E.R., Clifton,S.W., Chissoe,S.L., Marra,M.A., Raymond,C., Haugen,E.B., Gillett,W., Zhou,Y., James,R., Phelps,K., Iadamoto,S., Bubb,K., Simms,E., Levy,R., Clendenning,J., Kaul,R., Kent,W.J., Furey,T.S., Baertsch,R.A., Brent,M.R., Keibler,B., Flieck,P., Bork,P., Suyama,M., Bailey,J.A., Portnoy,M.E., Torrens,D., Chinwalla,A.T., Gish,W.R., Eddy,S.R., McPherson,J.D., Olson,M.V., Bichler,E.E., Green,E.D., Waterston,R.H. and Wilson,R.K. The DNA sequence of human chromosome 7 Nature 424 (6945), 157-164 (2003)
TITLE	Nature 424 (6945), 157-164 (2003)
JOURNAL	22737999
MEDLINE	12853948
PUBMED	
REFERENCE	2 (bases 1 to 190846)
AUTHORS	Hou,S., Small,J. and Wohldmann,P.
TITLE	The sequence of Homo sapiens BAC clone RP11-331D5
JOURNAL	Unpublished (2001)
REFERENCE	3 (bases 1 to 190846)
AUTHORS	Waterston,R.H.
TITLE	Direct Submission
JOURNAL	Submitted (11-JAN-1999) Genome Sequencing Center, Washington MO 63108, USA
REFERENCE	4 (bases 1 to 190846)
AUTHORS	Waterston,R.H.
TITLE	Direct Submission
JOURNAL	Submitted (05-MAY-1999) Genome Sequencing Center, Washington MO 63108, USA
REFERENCE	5 (bases 1 to 190846)
AUTHORS	Waterston,R.H.
TITLE	Direct Submission
JOURNAL	Submitted (02-MAY-2000) Genome Sequencing Center, Washington MO 63108, USA
REFERENCE	6 (bases 1 to 190846)
AUTHORS	Waterston,R.
TITLE	Direct Submission
JOURNAL	Submitted (30-SEP-2000) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE	7 (bases 1 to 190846)
AUTHORS	Wilson,R.
TITLE	Direct Submission
JOURNAL	Submitted (31-JAN-2004) Department of Genetics, Washington

COMMENT
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 On May 5, 1999 this sequence version replaced gi:1193346.
 ----- Genome Center
 Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: <http://genome.wustl.edu>
 Contact: sapiens@watson.wustl.edu
 ----- Summary Statistics
 Center project name: H_NH0331D05

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

ACKNOWLEDGMENTS:
The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/GT8/CHR7>, send <mailto:egreen@nhgri.nih.gov>, or see <http://genome.wustl.edu>

SOURCE INFORMATION:

The RPQC-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Woon, P. Y., Zhao, B., Frengen, E., Tatenio, M., Catanese, J. J., and de Jong, P. J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. *Genomics* 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Dieter de Jong and coworkers at <http://www.chori.org>

VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the right is RP4-703P8. Actual start of this clone is at base position 1 of RP11-331D5 actual end is at neighboring sequence information.

FEATURES

[illegible]

QY 120 ATTCTTTTCGATTAAATTAATCCAGATAGAGAGAGTAATTTTCGAAAGAAATGATA 179
|||||
Db 150898 ATTCAATTCGATTAAATTAATCCAGATAGAGAGAGTAATTTT-GGAAAGAAATGATA 150956
|||||
QY 180 GCTATATTAAACAGATATTCATTAACAATACCATGTAGAGACATAAGCAATATTTTGGCA 239
Db 150957 GCTATATTAAACAGATATTCATTAACAATACCATGTAGAGACATAAGCAATATTTTGGCA 151016
|||||
QY 240 TCAATTCGTCCCGTCAGTAGGCGGTGTTCCCTCTGGTAGGCGCTTGGAGATACCATCT 299
|||||
Db 151017 TCAATTCGTCCCGTCAGTAGGCGGTGTTCCCTCTGGTAGGCGCTTGGAGATACCATCT 151076
|||||
QY 300 ATCTAAGTAGGAGGAATGCTGTGGGAAGCGCGGATGGAGGTGCGTTTTCTACGCTGAAC 359
Db 151077 ATCTAAGTAGGAGGAATGCTGTGGGAAGCGCGGATGGAGGTGCGTTTTCTACGCTGAAC 151136
|||||
QY 360 CCCACACAGAAATCTGCAGCCACACAGCTGCCTCTGCGCGCCTTCATGTGATCATC 419
Db 151137 CCCACACAGAAATCTGCAGCCACACAGCTGCCTCTGCGCGCCTTCATGTGATCATC 151196
|||||
QY 420 CTGGTCATGAAGTGAATGTCCTATTTTCNGGGG 453
Db 151197 TTGGTCAATGAAGTGAATGTCCTATTTTCNGGG 151230
|||||

RESULT 10
BC034040/c 4735 bp mRNA linear PRI 30-JUN-2004
LOCUS BC034040
DEFINITION Homo sapiens protein tyrosine phosphatase, receptor type, N
polypeptide 2, transcript variant 3, mRNA (cDNA clone MGC:26506
IMAGE:4811578), complete cds.
ACCESSION BC034040
VERSION BC034040.1 GI:21707838
KEYWORDS MGC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 4735)
Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
Diachenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
Schettz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
Fahey,J., Helton,E., Kettman,M., Madan,A., Rodriguez,S.,
Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
Boutfard,G.G., Blakeley,R.W., Touchman,J.W., Green,E.D.,
Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
Butterfield,Y.S., Kraywinski,M.I., Skalska,U., Smalls,D.E.,
Scherch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
human and mouse cDNA sequences
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
12477932
2 (bases 1 to 4735)
Strausberg,R.
Direct Submission
Submitted (02-JUL-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgapps-r@mail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.

cDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki
Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome
Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: <http://www-shgc.stanford.edu>
Contact: (Dickson, Mark) mcd@paxil.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
R. M.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAK Plate: 32 Row: i Column: 20
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 19743910.

FEATURES

source

1..4735
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MGC:26506 IMAGE:4811578"
/tissue_type="Brain, hippocampus"
/clone_lib="NIH_MGC_95"
/lab_host="DH10B"
/note="Vector: pBluescript"

gene

1..4735
/gene="PTPRN2"
/note="synonyms: IA-2beta, IAR, ICAAR, KIAA0387, PTPRP,
IAR PTPRP"
/db_xref="LocusID:5799"
/db_xref="MIM:601698"
105..3065
/gene="PTPRN2"
/codon_start=1
/product="protein tyrosine phosphatase, receptor type, N
polypeptide 2, isoform 3 precursor"

CDS

105..3065
/db_xref="GI:21707839"
/db_xref="LocusID:5799"
/db_xref="MIM:601698"
/translation="MGPEPLPPLLLLLLLLPRVLPAAPSSVPRQLPQRLGLLLEEG
LCGSEACVNDGVFCRQKQVAMDFYEVSPVALQRLVQLKSLGTFGTQDDYDTQ
YMQDELADLPKLYLRREAPSPARPSKHSVGSRRYREGGAAALANLRHLPFLLEA
LSQAPSDVLARTHTAQRPAEGDDRFSESLTYVAHTSALTYPGPTQLREDDLLP
RTLGQPDLSKVDSPGVDHRLMAALSAVAORPPAPPGEGSLEPOVLLRAPSRMP
RELLAPAPQKWPSPGLDSEDPSSDGDARHTLLKDLQROPAEYRGLSLDGMAE
LMAGLMQGVHGVARGSPRALGSGSQADGPKATLGDSPFDGVDODDDRLYQEV
HRLSATLGLLQDHGSRLLPGLPFPARPLDMERKKSEHPSESLSSEETAGVENVKSQ
TYSKDLLGQHPSEFGAAAFGLQNMFGPKSEQSLPAGAQLSDGLQLEVPSEEE
EARGYIVTDREVILGPAVTFKVSANQVNTDEVKATVDNKKLEETSGKLITQTVG
SKSKLFLPPQAEQSDSTKFIALTJVLACILGVLLASGLIYCLRHSQRLKELSG
LGGDPGADATAYQELCRQMATRPDRPEGPTSRISVSQSDGPTSPSARSAA
SSWSEEPQSNMDISTGHMILSYMEDHLKNKLEKEALCAYQAEPNSSFVAORSE
NVPKNRSLAVLTYDHSRVLKKAENSHSDYINASPIHMDPRNPAYIATQPLPATV
LVRFWLMQNTQETRTVTPLAENGVRQCYHYWPDGNSLHYIYENVLSEHICEDF
ADRSYKLNQNTETRTVTPHFLSWYDRGVSPSSRLDFFRKKNKCYGRGSCPTIV
HCSDCAGSGTIVLDMVLNKMKAKEIDIAATLEHLRDRQRPQGVQTEQEFALTA
VAEYVNAIKALPQ"

ORIGIN

Query Match 91.4%; Score 416; DB 9; Length 4735;
Best Local Similarity 98.9%; Pred. No. 1,1e-92;
Matches 449; Conservative 0; Mismatches 2; Indels 3; Gaps 3;
QY 1 GAAATCAGAGTTTAAATATGACACAAATTAATATTTTGTATATCTCACACCGGA-GNNT 59
|||||
Db 4716 GAAATCAGAGTTTAAATATGACACAAATTAATATTTTGTATATCTCACACCGAGTTT 4657
|||||
QY 60 CTCTTCAACATAAGGAGTTAGAAATTACAGTAGGCATATGCTTCTTATATTCAGATAA 119
|||||
Db 4656 CTCTTCAACATAAGGAGTTAGAAATTACAGTAGGCATATGCTTCTTATATTCAGATAA 4597
|||||
QY 120 ATTCAATTCGATTAATAATTAATTCAGATAGAGAGAGTAATTTTCGGAAGAAAGATGATA 179
|||||

```

4596 ATTCAATTCGATTAAATAATCCAGATAGAGAGAAATAATTTT-GGAAAAAGAAATGATA 4538
180 GCTATATTAAAGCAGATATTTCATTACAATACCATGTAGACATAGAAGCAATATTTTGGCA 239
4537 GCTATATTAAAGCAGATATTTCATTACAATACCATGTAGACATAGAAGCAATATTTTGGCA 4478
240 TCATTCCTCCGCTCAGTAGGCGGTGTTCCCTCTGTTAGGCGCTTTGGAGAGTACCATCT 299
4477 TCATTCCTCCGCTCAGTAGGCGGTGTTCCCTCTGTTAGGCGCTTT-GAGAGTACCATCT 4419
300 ATCTAAGATGAGGAATCTGTGGAGGCGGATGAGGTGCGTTTCTACGCTGAAC 359
4418 ATCTAAGATGAGGAATCTGTGGAGGCGGATGAGGTGCGTTTCTACGCTGAAC 4359
360 CCCACACAGGAATCTGCAGGCCACACACAGCTGCTCTGCGCGCTTCCATCTGATCATC 419
4358 CCCACACAGGAATCTGCAGGCCACACACAGCTGCTCTGCGCGCTTCCATCTGATCATC 4299
420 CTGGTCAATGAAGTGAATTCCTATTTTCNGGGG 453
4298 CTGGTCAATGAAGTGAATTCCTATTTCTGGG 4265

RESULT 11
HSU66702/c 4719 bp mRNA linear PRI 18-OCT-1996
LOCUS Human phogrin mRNA, complete cds.
DEFINITION
ACCESSION U66702
VERSION U66702.1 GI:1620663
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 4719)
AUTHORS Kawasaki, E., Hutton, J.C. and Eisenbarth, G.S.
TITLE Molecular cloning and characterization of the human transmembrane
protein tyrosine phosphatase homologue, phogrin, an autoantigen of
type 1 diabetes
JOURNAL Biochem. Biophys. Res. Commun. 227 (2), 440-447 (1996)
MEDLINE 97032784
PUBMED 8878534
REFERENCE
2 (bases 1 to 4719)
AUTHORS Kawasaki, E., Hutton, J.C. and Eisenbarth, G.S.
TITLE Direct Submission
JOURNAL Submitted (12-AUG-1996) Barbara Davis Center for Childhood
Diabetes, University of Colorado Health Sciences Center, 4200 East
9th Avenue, Box B-140, Denver, CO 80262, USA
FEATURES
source
1..4719
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/cell_type="islet"
/tissue_type="pancreas"
15..3052
/note="protein tyrosine phosphatase homolog"
/codon_start=1
/product="phogrin"
/protein_id="AAC50742.1"
/db_xref="GI:1620664"
/translation="MGPPPLPILLLLLLPPRVLPAAPSSVPRGRQLPRLGCLLEEG
LCGASEACVNDGVFCOKPAMDVRYEVSVVALQRLVALQKLSGTGFTWDDYTO
LYMDELADLPKTYLRRPEASSPARPSKHSVSRYSREGGAALANALRHLPLEA
LSQAPASDLAKHTAQDRPPAGDRFSILITVAHTSALTYPGPRQTQREDLLP
RTLQQLQPELSPKVSDSGVDRLHMAALSAYAAQRPAPPGEGSLEPQYLRAPSRMP
RPLAPAPQKWPSPSGSDPSSTGDSGARITHLKDLQRPFAEVRGLSGLELDGMAE
LMAGLMQVDHGVARGSPRAALGSDGPKATLRGDFPDGVDQDDRLYQEV
HRLSATLGLGLDHGSRLLPGALPFARPLDMERKKEHPESLSSEETAGVENVKSO
TYSKDLGQQHPSEGAALFGLQNMFGPKSEQSLFAGAEALSDGLQLEVPSEE
EARGYIVTRDLPREGRRLVEDVARLLQVPSSAFADVGLGPAVITFKVSANQVNT
TEDVEKATVDNKKLEETSGLKILQTVGSGSKSLKFLPPOAEQEDSTKFIATLVSLA

```

```

CILGVLLASGLIYCLRHSSOHLRKEKLSGLGDPGADATAYQELCRQMATRPDRP
EGPHTSISSVSOFSDGPIPSPSARSASSNSWSEPVQSNMDISTGHMILSYMEDHLX
NKRLEKEWALCAIYQAEPNSSFVAQREENVPKNRSIAVLITYDHSRVLKKAENSHLS
DYNASIPMDHPENPAYIATQGPLPATVADFWMQWESGCVIIMLTPLAENGVRQC
YHYWPDGSLNLIHYEVLNVEHIFWCBDFLVRFSYFLKNLQTNETRTVTQPHLSWYDR
GVPSSSSLADFRKVNKCYGRSCPIIIVHSCDGAGRSCTVVLIDMLNWKAKGAEI
DIATLEHLRDRQRFQVTKEQPFALTAVAEVNAIILKALPQ"

ORIGIN
Query Match 91.1%; Score 414.6; DB 9; Length 4719;
Best Local Similarity 98.2%; Pred. No. 2.5e-92;
Matches 439; Conservative 0; Mismatches 6; Indels 2; Gaps 2;

QY 8 GAAGTTTAATATGACACAAATTAATATATTTGTATATCTCACCCGGA-GNVTCTCTTCA 66
Db 4710 GAAGTTTAATATGACACAAATTAATATATTTGTATATCTCACCCGGAAGTTTCTCTTCA 4651
QY 67 AACATAAGGAGTGTAGAATTAACAAGTAGGCAATATGCTTCTTATATTCAGATAAAATTCATT 126
Db 4650 AACATAAGTGTGTAGAATTAACAAGTAGGCAATATGCTTCTTATATTCAGATAAAATTCATT 4591
QY 127 TCGATTAAATTAATTCAGATAGAGAGAAATAATTTTCGAAAAAGAAATGATAGTATAT 186
Db 4590 TCGATTAAATTAATTCAGATAGAGAGAAATAATTTTCGAAAAAGAAATGATAGTATAT 4532
QY 187 TAAAGCAGATATTCATTACAATACCATGTAGACATAGAACAATATTTTGGCATCATCT 246
Db 4531 TAAAGCAGATATTCATTACAATACCATGTAGACATAGAACAATATTTTGGCATCATCT 4472
QY 247 GTCCGCTCAGTAGGCGGTGTTCCCTCTGTTAGGCGCTTTGGAGAGTACCATCTATTAAG 306
Db 4471 GTCCGCTCAGTAGGCGGTGTTCCCTCTGTTAGGCGCTTTGGAGAGTACCATCTATTAAG 4412
QY 307 ATGAGGAATCTCTGGAAGCGGGATGAGGTGCGTTTTTCTACGCTGAACCCACAC 366
Db 4411 ATGAGGAATCTCTGGAAGCGGGATGAGGTGCGTTTTTCTACGCTGAACCCACAC 4352
QY 367 AGGAAATCTGAGCCACACAGCTGCTCTGCGCGCTTCCATGTGATCATCTCTGGTCA 426
Db 4351 AGGAAATCTGAGCCACACAGCTGCTCTGCGCGCTTCCATGTGATCATCTCTGGTCA 4292
QY 427 ATGAAGTGAATTCCTATTTTCNGGG 453
Db 4291 ATGAAGTGAATTCCTATTTTCNGGG 4265

RESULT 12
GI9707 human STS A001W14, sequence tagged site. STS 24-JUL-1996
LOCUS
DEFINITION
ACCESSION GI9707
KEYWORDS STS; STS sequence; primer; sequence tagged site.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 105)
Adams M.D.
Unpublished (1996)
CONTACT
Contact: Mark Adams
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850
Email: mdadams@tigr.org
PRIMER
Primer A: ACAATACCATGTAGACATAGAAG
Primer B: ATCTTAGATAGTGGTACTCTCC
STS size: 105
PCR Profile:
Denaturation: 96C 5min
Anneal: 54C 30sec
Extend: 72C 30sec

```

Denature: 95C 30sec
FinalExtend: 72C 5min
Cycles: 30

Protocol:

GenomicDNA: 25 ng
Primer: 0.43 uM each
dNTPs: 230 uM each
Amplify: 0.5 units
TagStart Ab: 0.5 units
Total Volume: 10 ul

Buffer:

Tris-HCl pH8.8: 100 mM
KCl: 500 mM
MgCl2: 20 mM
Triton X-100: 1%
Concentration: 10X

Prepared with primer pairs derived from THCS1137; GenBank Accession Numbers-- F09140, F10434, T03688, T15930, T19338, T32236, T33356, T33873.

FEATURES

source

1..105
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
1..105
1..23
complement (83..105)

STS

primer_bind
primer_bind

ORIGIN

Query Match 23.1%; Score 105; DB 11; Length 105;
Best Local Similarity 100.0%; Pred. No. 1.6e-15;
Matches 105; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 204 ACAATACCATGTAGACATACCAATATTTGGCATCATCTTGTCCGCTCAGTAGCGG 263
|||||
Db 1 ACAATACCATGTAGACATACCAATATTTGGCATCATCTTGTCCGCTCAGTAGCGG 60
|||||

QY 264 TGTTCCTCTGGTAGGCGCTTTGGAGAGTACCATCTATCTAGAT 308
|||||
Db 61 TGTTCCTCTGGTAGGCGCTTTGGAGAGTACCATCTATCTAGAT 105
|||||

RESULT 13

AC117965/c

LOCUS

DEFINITION

Rattus norvegicus clone CH230-366B11, *** SEQUENCING IN PROGRESS

***, 2 unordered pieces.

AC117965

GI:25138129

HTG: HTGS_PHASE1; HTGS DRAFT: HTGS_ENRICHED.

KEYWORDS

Rattus norvegicus (Norway rat)

SOURCE

ORGANISM

Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 180187)

REFERENCE

AUTHORS

Muzny,D.Marie., Metzker,M.Lee., Abramson,S., Adams,C., Alder,J., Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D., Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Caesar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,

Gebregorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpathy,S., Kelly,S., King,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensuhewa,L., Loulseghe,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindaratne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mawhney,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Parks,K., Nwaokeme,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Prannkoch,C., Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savary,G., Scherer,S., Scott,C., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajls,D., Speed,A., Sodergren,B., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,N., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Umani,K., Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K., Wright,D., Wright,J., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,X., Zhou,S., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.

Direct Submission

Unpublished

2 (bases 1 to 180187)

Worley,K.C.

Direct Submission

Submitted (12-APR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 180187)

Rat Genome Sequencing Consortium.

Direct Submission

Submitted (20-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On Nov 20, 2002 this sequence version replaced gi:23811557.

The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: http://www.hgsc.bcm.tmc.edu/

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GUOM

Center clone name: CH230-366B11

----- Summary Statistics

Assembly program: Phrap; version 0.990329

Consensus quality: 169631 bases at least Q40

Consensus quality: 171486 bases at least Q30

Consensus quality: 172615 bases at least Q20
 Estimated insert size: 175046; sum-of-contigs estimation
 Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

*** NOTE: Estimated insert size may differ from sequence length
 (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * NOTE: this is a 'working draft' sequence. It currently
 * consists of 2 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

* 1 178750: contig of 178750 bp in length
 * 178751 178850: gap of unknown length
 * 178851 180187: contig of 1337 bp in length.

FEATURES
 source
 1..180187
 /organism="Rattus norvegicus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10116"
 /clone="CH230-366811"
 1..1502
 /note="wgs_contig"
 172128..173041
 /note="clone_boundary"
 clone_end:17
 site:
 end_sequence:B2121917"
 176880..178750
 /note="wgs_end_extension"
 clone_end:17"

misc_feature
 18.9%; Score 86.2; DB 2; Length 180187;
 Best Local Similarity 71.5%; Pred. No. 1.2e-10;
 Matches 143; Conservative 0; Mismatches 49; Indels 8; Gaps 2;

misc_feature
 172128..173041

misc_feature
 176880..178750

ORIGIN
 Query Match 18.9%; Score 86.2; DB 2; Length 180187;
 Best Local Similarity 71.5%; Pred. No. 1.2e-10;
 Matches 143; Conservative 0; Mismatches 49; Indels 8; Gaps 2;

Qy 9 AAGTTTAAATGACACAAATTAAATATATCTGTATATCTCACCGGAGNTTCTCTTCAA 68
 Db 163149 AGTTTAAATAGACACAAATTAAATATATCTGTATATCTCACCGGAGNTTCTCTTCAA 163090

Qy 69 CATAAGGAGTTAGAAATACAAATGAGCATATGCTTCTTATATTCAGATAAATTCATTT 128
 Db 163089 CATAAGAG-----TTACAATTAAGTATATGCTTCTTATATTCAGATAAATTCATTT 163037

Qy 129 GATTAAATTAATTCAGATAGAGAGTAATTTTCGGAAGAAATGATAGCTATATTA 188
 Db 163036 TATTAATTAATTCATTTAGATTCGGAAGGGAATGGAAGGAACG-TCGCAGATGTT 162978

Qy 189 AAGCAGATATTCATTACCAAT 208
 Db 162977 AAGCAGATACACACAAAT 162958

RESULT 14
 RN073458/c 5136 bp mRNA linear ROD 26-MAR-1998
 LOCUS Rattus norvegicus protein tyrosine phosphatase (PTPNE6) mRNA,
 DEFINITION complete cds.
 U73458
 ACCESSION U73458.1 GI:1657945
 VERSION
 KEYWORDS Rattus norvegicus (Norway rat)
 SOURCE Rattus norvegicus
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 5136)
 REFERENCE Fitzgerald,L.R., Walton,K.M., Dixon,J.E. and Largent,B.L.
 AUTHORS PTP NE-6: a brain-enriched receptor-type protein tyrosine
 TITLE phosphatase with a divergent catalytic domain

J. Neurochem. 68 (5), 1820-1829 (1997)
 97263591
 MEDLINE
 PUBMED 9109506
 REFERENCE 2 (bases 1 to 5136)
 AUTHORS Fitzgerald,L.R., Walton,K.M., Dixon,J.E. and Largent,B.L.
 TITLE Direct Submission
 JOURNAL Submitted (04-OCT-1996) CNS Diseases Research, DuPont Merck
 Research Labs, Experimental Station E400, Wilmington, DE 19880, USA
 FEATURES
 source
 1..5136
 /organism="Rattus norvegicus"
 /mol_type="mRNA"
 /strain="Sprague-Dawley"
 /db_xref="taxon:10116"
 1..5136
 /gene="PTPNE6"
 112..3126
 /gene="PTPNE6"
 /function="receptor-type"
 /codon_start=1
 /product="protein tyrosine phosphatase"
 /protein_id="AAC08036.1"
 /db_xref="GI:1657945"
 /translation="MGLPLPLLLPLPPALPAPASARGQLPRLGLCFEDGL
 CGSLETGVNDGVFGRCKVPALDTYRYEVSFGALLHRIILQKLSRTGFTWDDYTQR
 VIAQLSNLPKAYLWHEEASPARSLQONADNEKWFSEVALAKTRRYLPYLELL
 SQAPTAHNPRIIDHETRVKGEDSPENILTYAHTSALTYPPTATKRVYDNLRLPLS
 RLQDELSPKVDSDIDKQKLIAGAYTAQPPGENDPEPRYLVHSPMRADSPFAAPA
 LSORWPLPGDSKDSLMDGDTLLRSLLKDLQQAEDVRLGSLKLEQADSIAAGIAS
 DPVEGSESHGRGAEGLREQADAPBEMLQDRLPEVDDPAAYKEVRSYKDLLRKPN
 DHGSPLLPEAPLLEKSRAMKSEQPEEVLSEETAGVEHVKSYSKDLRLKPN
 SEFPWRLEDFQNPFAVEWEDQNLAAQSPSGGLQLEVQSEEEQQQYIITGN
 PLSPEKQKLMDEVAHLRVPFSDVAVKLVGPAVIFKVSANIQMTTADVTKAAND
 KBLEKATGLTILQSGIRPKGKLLPHEEEDSTFKVILFSLIACILAVLLASSL
 AYCLRNSHYKLKELGSLGADPSADATEAVOELCRORMAVRPDGHSEGHPTSRNSV
 SSQSDGPMPSPARSTSSWSBEPASQNDISTGHMLAYMEHDLKMKNELEWEA
 LCATQBPDSLSVAQREENAPKRSALVITDHSRILLKSESHNSNDYINASPMIDH
 DPNPAYIATQGPLPATVADFQWVMVSGCAIVMLTPLSENGVRQCHHYMPDGSNV
 YHYEVNLSVEHIWCQDFLVRSYKLNQTNRTVTQPHFLSWYDQGVSPSTSLLD
 FRKVNKYGRSCPIIVHCSGDAGRSCTVLIDMVLNKAAGKAKEIDIAATLEHLRD
 QRPQWQTKQEFALTAVAEVNAILKALPQ"

ORIGIN
 Query Match 18.8%; Score 85.6; DB 10; Length 5136;
 Best Local Similarity 71.3%; Pred. No. 1.4e-10;
 Matches 129; Conservative 0; Mismatches 45; Indels 7; Gaps 1;

Qy 9 AAGTTTAAATGACACAAATTAAATATATTTCTATATCTCACCGGAGNTTCTCTTCAA 68
 Db 4567 AGGTTTAAATAGGACACAAATTAAATATATCTGTATATCACCGGAGGTTTTCTCTTAAA 4508

Qy 69 CATAAGGAGTTAGAAATACAAATGAGCATATGCTTCTTATATTCAGATAAATTCATTT 128
 Db 4507 CATAAGAG-----TTACAATTAAGTATATGCTTCTTATTCAGATAAATTCATTT 4455

Qy 129 GATTAAATTAATTCAGATAGAGAGTAATTTTCGGAAGAAATGATAGCTATATTA 188
 Db 4454 TATTAATTAATTCATTTAGATTCGGAAGGGAATGGAAGGAACGCTCCAGATGTTA 4395

Qy 189 A 189
 Db 4394 A 4394

RESULT 15
 AC110936
 LOCUS Rattus norvegicus clone CH230-188N19, *** SEQUENCING IN PROGRESS
 DEFINITION ***; 4 unordered pieces.
 ACCESSION AC110936
 VERSION AC110936.6 GI:25072900
 KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
 SOURCE Rattus norvegicus (Norway rat)

ORGANISM

Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE
AUTHORS

1 (bases 1 to 222251)
Muzny, D., Marie, Metzker, M., Lee, A., Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C. A., Fallis, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K., Harvey, Y., Haylak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowig, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorenshewa, L., Loulseghe, H., Lozado, R. J., Lu, X., Ma, J., Maheshwari, M., Mahindaratne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapa, P., Martin, K., Martin, R., Martinez, E., Mawhiney, S., McLeod, M. P., McNeill, T. Z., Meenen, E., Mileavljivic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Naif, L., Nankervils, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwaokeme, O., Okunolu, G., Olarnpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L., Pu, L., Pu, M., Quirroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J., Sanders, W., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbay, A., Sisson, I., Sitter, C. D., Smajs, D., Speed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steidle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valae, R., Vera, V., Villabana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, P., Williams, G., Willson, R., Wleczyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausen, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O., Weinstock, G., and Gibbs, R. A.

TITLE
JOURNAL

Direct Submission
2 (bases 1 to 222251)
Worley, K.C.
Direct Submission
Submitted (17-FEB-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 222251)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (19-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE
AUTHORS

1 (bases 1 to 222251)
Submitted (17-FEB-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 222251)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (19-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

COMMENT

On Nov 19, 2002 this sequence version replaced gi:23270153.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas

assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GOFU
Center clone name: CH230-188N19
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 193700 bases at least Q40
Consensus quality: 196702 bases at least Q30
Consensus quality: 198783 bases at least Q20
Estimated insert size: 196105; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 213437: contig of 213437 bp in length
* 213438 213537: gap of unknown length
* 213538 218989: contig of 5452 bp in length
* 218990 219089: gap of unknown length
* 219090 220660: contig of 1571 bp in length
* 220661 220760: gap of unknown length
* 220761 222251: contig of 1491 bp in length.
* Location/Qualifiers

FEATURES
source

1..222251
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-188N19"

misc_feature

1..1121
/note="wgs_contig"

misc_feature

16434..18286
/note="wgs_contig"

misc_feature

27659..28735
/note="wgs_contig"

misc_feature

128775..130020
/note="wgs_contig"

misc_feature

174053..175360
/note="wgs_contig"

misc_feature

183232..184558
/note="wgs_contig"

misc_feature

213538..215409
/note="wgs_contig"

ORIGIN

Query Match 16.9%; Score 76.8; DB 2; Length 222251;
Best Local Similarity 80.6%; Pred. No. 2.7e-08;
Matches 104; Conservative 0; Mismatches 18; Indels 7; Gaps 1;

QY 9 AAGTTTAATATACACACAAATTAATATATTTGATATCTCACACCGAGNTTCTCTTCAA 68

Db 127895 AGTTTAAATAGACACAAATTAATATCTGATATATCACACCGAGGTTTTCCTTTAA 127954

QY 69 CATAGAGATTGAGAANTTCAAGTAGGCGATATGCTTCCTATATTAGATAAATTCAATTC 128

Db 127955 CATAAGAAG-----TTACAAATTAAGTATATGCTTCCTATATTCAGATAAATTCATTTC 128007

Qy 129 GATTAATTA 137

Db 128008 TATTAATTA 128016
|||||

Search completed: May 5, 2005, 03:58:33
Job time : 2231.09 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 4, 2005, 20:02:41 ; Search time 314.957 Seconds
(without alignment)
8551.893 Million cell updates/sec

Title: US-08-731-499-8
Perfect score: 455
Sequence: 1 GAATCAGAACTTTAATG.....ATTGCTCTATTTCNGGGGGT 455

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_16Dec04.*
1: Geneseqn1980s.*
2: Geneseqn1990s.*
3: Geneseqn2000s.*
4: Geneseqn2001as.*
5: Geneseqn2001bs.*
6: Geneseqn2002as.*
7: Geneseqn2002bs.*
8: Geneseqn2003as.*
9: Geneseqn2003bs.*
10: Geneseqn2003cs.*
11: Geneseqn2003ds.*
12: Geneseqn2004as.*
13: Geneseqn2004bs.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	453	99.6	455	AAV09022	Aav09022 Homo sapi
C 2	439	96.5	1258	Aah34753	Aah34753 Human col
C 3	426.4	93.7	2328	Aat95227	Aat95227 Human isl
C 4	426	93.6	4699	Adl14680	Adl14680 Human src
C 5	426	93.6	4767	Adn95437	Adn95437 Human BEC
C 6	414.6	91.1	4719	Adp18660	Adp18660 Human TAR
C 7	254.8	56.0	282	Abv95893	Abv95893 Human pan
C 8	225.2	49.5	4856	Aas87633	Aas87633 DNA encod
9	120.6	26.5	121	Adf86738	Adf86738 Single nu
10	114.6	25.2	121	Adf86737	Adf86737 Single nu
11	109.6	24.1	121	Adf86735	Adf86735 Single nu
12	107.6	23.6	121	Adf86734	Adf86734 Single nu
13	92.6	20.4	121	Adf86736	Adf86736 Single nu
C 14	64.4	14.2	455	Aah35240	Aah35240 Human col
C 15	54.6	12.0	538	Adq53174	Adq53174 Novel can
16	49.4	10.9	51	Aal29285	Aal29285 Human SNP
C 17	45.4	10.0	2000	Ada71938	Ada71938 Rice gene
18	44.2	9.7	6811	Abi34540	Abi34540 Human met
19	44.2	9.7	6811	Abi70263	Abi70263 Chemical
20	44.2	9.7	6811	AdS99801	AdS99801 Bisulphit

C 21	44.2	9.7	11422	6	ABK39937	Abk39937 Human che
C 22	44.2	9.7	11422	6	ABL32219	Abi32219 Human imm
C 23	44.2	9.7	37515	6	ABQ66997	Abq66997 Human ang
24	44	9.7	2000	8	ADA71938	Ada71938 Rice gene
25	43	9.5	1246	2	AZ42154	Aaz42154 Human nor
26	43	9.5	2117	2	AAT61258	Aat61258 Human cdn
27	43	9.5	2815	3	AAF21830	Aaf21830 Human bre
28	43	9.5	3323	13	ADR67210	Adr67210 Human bla
29	43	9.5	3323	13	ADR66257	Adr66257 Human pro
30	43	9.5	3323	13	ADR66599	Adr66599 Human pro
C 31	43	9.5	61020	4	AA546787	Aas46787 Tumour su
32	42.6	9.4	51	4	AAL29288	Aal29288 Human SNP
33	42.6	9.4	344	4	AAH93307	Aah93307 Human SNP
34	42	9.2	51	4	AAAL29282	Aal29282 Human SNP
35	42	9.2	1132	3	AAF12929	Aaf12929 Aspergill
C 36	42	9.2	8056	8	ABZ10246	Abz10246 Haematopo
C 37	42	9.2	13574	6	ABL33317	Abi33317 Human imm
C 38	41.6	9.1	2849	4	ABL14726	Abi14726 Drosophil
C 39	41.4	9.1	110000	13	ABD32966_09	Continuation (10 o
C 40	41.2	9.1	6988	6	ABL34441	Abi34441 Human imm
41	41	9.0	11922	3	AAA70187	Aaa70187 Plasmodiu
C 42	40.6	8.9	794	6	ABQ42952	Abq42952 Oligonuc1
43	40.6	8.9	794	6	ABQ42953	Abq42953 Oligonuc1
C 44	40.4	8.9	337	12	ADO55043	Ado55043 Gene #140
45	40.4	8.9	700	4	AAH93026	Aah93026 Human inf

ALIGNMENTS

RESULT 1
AAV09022
ID AAV09022 standard; cDNA; 455 BP.
XX
AC AAV09022;
XX
DT 21-JUL-1998 (first entry)
XX
DE Homo sapiens 20ql3 amplicon 20ea7 transcript.

XX 20ql3 amplicon; chromosome 20; tumour; detection;
KW chromosomal abnormalities; probe; gene therapy; antisense inhibition;
KW treatment; age-related macular degeneration; retinitis pigmentation;
KW Leber's congenital amaurosis; BEM-1; ds.
XX Homo sapiens.
XX WO9802539-A1.
XX
PD 22-JAN-1998.
XX
PF 15-JUL-1997; 97WO-US012343..
XX
PR 15-JUL-1996; 96US-00680395.
PR 16-OCT-1996; 96US-00731499.
PR 17-JAN-1997; 97US-00785532.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Gray JW, Collins CC, Hwang S, Godfrey T, Kowbel D, Rommens J;
XX WPI; 1998-110587/10.
XX
DR New sequences from the 20ql3 amplicon - used for detecting chromosomal
XX abnormalities, particularly tumours, and for developing products for
PT treating diseases.
PT
XX Claim 1; Page 63-64; 91pp; English.
XX
CC The sequence is that of a cDNA sequence 20ea7 which was isolated from the
CC 20ql3 amplicon. It is a homologue of the rat gene BEM-1 and can be used
CC as a probe for the detection of chromosomal abnormalities at 20ql3. It
CC and other sequences isolated from the 20ql3 amplicon are consistently

CC amplified in primary tumours. These sequences are useful as probes or as
CC probe targets for monitoring the relative copy number of corresponding
CC sequences from a biological sample such as tumour cells. The sequences
CC can also be used in therapeutic applications for modulating the
CC expression of the endogenous gene or the activity of the gene product.
CC Examples of therapeutic approaches include antisense inhibition of gene
CC expression, gene therapy, and monoclonal antibodies that specifically
CC bind the gene products. The products can also be used in the treatment of
CC other diseases, e.g. age-related macular degeneration, Leber's congenital
CC amaurosis and retinitis pigmentata
XX
SQ Sequence 455 BP; 138 A; 87 C; 98 G; 130 T; 0 U; 2 Other;

Query Match 99.6%; Score 453; DB 2; Length 455;
Best Local Similarity 100.0%; Pred. No. 5e-118;
Matches 455; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GAAATCAGAGTTTAATATGACACAATTAATATATTTGTATATCTCACACGGAGNTTC 60
DB 1 GAAATCAGAGTTTAATATGACACAATTAATATATTTGTATATCTCACACGGAGNTTC 60
QY 61 TCTTCAACATAGGAGTTAGAAATACAGTAGGAGATATGCTTCTATATTCAGATAAA 120
DB 61 TCTTCAACATAGGAGTTAGAAATACAGTAGGAGATATGCTTCTATATTCAGATAAA 120
QY 121 TTCAATTTCCGATTAAATTAATTCAGATAGAGAGAAGTAATTTTCGGAAGAAGAAATCATAG 180
DB 121 TTCAATTTCCGATTAAATTAATTCAGATAGAGAGAAGTAATTTTCGGAAGAAGAAATCATAG 180
QY 181 CTATATTAAGCAGATATTCATTTACAAATACCATGTAGAGACATAPAGCAATATTTTGGCAT 240
DB 181 CTATATTAAGCAGATATTCATTTACAAATACCATGTAGAGACATAPAGCAATATTTTGGCAT 240
QY 241 CATTCGTCCGCTCAGTAGGCCGTGTTCCCTCTGTTAGGAGGCTTTGGAGATACCATCTA 300
DB 241 CATTCGTCCGCTCAGTAGGCCGTGTTCCCTCTGTTAGGAGGCTTTGGAGATACCATCTA 300
QY 301 TCTAAGATGGAGGAATGCTGTGGGAAGGGGGGATGGAGTGCGTTTCTACGCTGAACC 360
DB 301 TCTAAGATGGAGGAATGCTGTGGGAAGGGGGGATGGAGTGCGTTTCTACGCTGAACC 360
QY 361 CCACACAGGAAATCTGAGCCCAACACAGCTGCTCTGCGCGGCTTCCATGTGATCATCC 420
DB 361 CCACACAGGAAATCTGAGCCCAACACAGCTGCTCTGCGCGGCTTCCATGTGATCATCC 420
QY 421 TGGTCAATGAAGTGAATTCCTATTTTCNGGGGT 455
DB 421 TGGTCAATGAAGTGAATTCCTATTTTCNGGGGT 455

RESULT 2
AAH34753/c
ID AAH34753 standard; cDNA; 1258 BP.
XX
AC AAH34753;
XX
AT 03-SEP-2001 (first entry)
XX
DE Human colon cancer antigen encoding cDNA SEQ ID NO:1835.
XX
KW Human; colon cancer; colon cancer antigen; diagnosis; detection;
XX colorectal carcinoma; chromosome 7; ss.
OS Homo sapiens.
XX
PN WO200122920-A2.
XX
PD 05-APR-2001.
XX
PF 28-SEP-2000; 2000WO-US026524.
XX
PR 29-SEP-1999; 99US-0157137P.
PR 03-NOV-1999; 99US-0163280P.

(HUMA-) HUMAN GENOME SCI INC.
PI Ruben SM, Barash SC, Birse CE, Rosen CA;
XX
DR WPI; 2001-235357/24.
DR P-PSDB; AAG75348.
XX
PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
PT useful for preventing, diagnosing and/or treating colorectal cancers.
XX
PS Claim 1; Page 3372-3373; 9803pp; English.
XX
CC AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
CC cancer-associated nucleic acid molecules (N) and proteins (P), where the
CC proteins are collectively known as colon cancer antigens. The colon
CC cancer antigens have cytostatic activity and can be used in gene therapy
CC and vaccine production. N and P may be used in the prevention, diagnosis
CC and treatment of diseases associated with inappropriate P expression. For
CC example, N and P may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of P by expressing inactive proteins or to
CC supplement the patients own production of P. Additionally, N may be used
CC to produce the colon cancer-associated Ps, by inserting the nucleic acids
CC into a host cell and culturing the cell to express the proteins. N and P
CC can be used in the prevention, diagnosis and treatment of colorectal
CC carcinomas and cancers. AAH37196 to AAH37204 and AAG77789 represent
CC sequences used in the exemplification of the present invention. N.B.
CC Pages 666 to 682 and page 7053 of the sequence listing were missing at
CC time of publication, meaning no sequences are present for SEQ ID NO:1027
CC to 1052, 7921 and 7922
XX
SQ Sequence 1258 BP; 345 A; 325 C; 302 G; 284 T; 0 U; 2 Other;
Query Match 96.5%; Score 439; DB 4; Length 1258;
Best Local Similarity 99.3%; Pred. No. 6.9e-114;
Matches 450; Conservative 0; Mismatches 2; Indels 1; Gaps 1;
QY 1 GAAATCAGAGTTTAATATGACACAATTAATATATTTGTATATCTCACACGGAGNTTC 60
DB 1213 GAAATCAGAGTTTAATATGACACAATTAATATATTTGTATATCTCACACGGAGNTTC 1154
QY 61 TCTTCAACATAGGAGTTAGAAATTAAGTAGGATATGCTTCTATATTCAGATAAA 120
DB 1153 TCTTCAACATAGGAGTTAGAAATTAAGTAGGATATGCTTCTATATTCAGATAAA 1094
QY 121 TTCAATTTCCGATTAAATTAATTCAGATAGAGAGAAGTAATTTTCGGAAGAAGAAATCATAG 180
DB 1093 TTCAATTTCCGATTAAATTAATTCAGATAGAGAGAAGTAATTTTCGGAAGAAGAAATCATAG 1035
QY 181 CTATATTAAGCAGATATTCATTTACAAATACCATGTAGAGACATAPAGCAATATTTTGGCAT 240
DB 1034 CTATATTAAGCAGATATTCATTTACAAATACCATGTAGAGACATAPAGCAATATTTTGGCAT 975
QY 241 CATTCGTCCGCTCAGTAGGCCGTGTTCCCTCTGTTAGGAGGCTTTGGAGATACCATCTA 300
DB 974 CATTCGTCCGCTCAGTAGGCCGTGTTCCCTCTGTTAGGAGGCTTTGGAGATACCATCTA 915
QY 301 TCTAAGATGGAGGAATGCTGTGGGAAGGGGGGATGGAGTGCGTTTCTACGCTGAACC 360
DB 914 TCTAAGATGGAGGAATGCTGTGGGAAGGGGGGATGGAGTGCGTTTCTACGCTGAACC 855
QY 361 CCACACAGGAAATCTGAGCCCAACACAGCTGCTCTGCGCGGCTTCCATGTGATCATCC 420
DB 854 CCACACAGGAAATCTGAGCCCAACACAGCTGCTCTGCGCGGCTTCCATGTGATCATCC 795
QY 421 TGGTCAATGAAGTGAATTCCTATTTTCNGGGG 453
DB 794 TGGTCAATGAAGTGAATTCCTATTTTCNGGGG 762
RESULT 3
AAT95227/c

ID AAT95227 standard; cDNA; 2328 BP.
XX AAT95227;
AC
XX 27-MAR-1998 (first entry)
DT
XX Human islet cell antigen 1851 partial cDNA clone WK121315.
DE
XX Islet cell antigen 1851; autoantigen; protein tyrosine phosphatase;
KW insulin-dependent diabetes mellitus; IDDM; diagnosis; therapy; human; ds.
XX Homo sapiens.
OS
XX WO9732984-A1.
PN
XX 12-SEP-1997.
PD
XX 05-MAR-1997; 97WO-US003532.
PF
XX 06-MAR-1996; 96US-0012927P.
PR 15-OCT-1996; 96US-0027540P.
XX (ZYMO) ZYMOGENETICS INC.
PA (UNIV) UNIV WASHINGTON.
XX Kindsvogel W, Jelinek L, Sheppard PO, Hagopian W, Lagasse J;
PI WPI; 1997-457535/42.
DR Mammalian pancreatic islet cell antigen and related DNA - used to detect
PT auto-antibodies indicative of insulin-dependent diabetes or pre-
PT disposition to it.
XX Example 5; Page 114-115; 134pp; English.
PS
XX This cDNA sequence comprises clone WK121315 a partial cDNA clone of novel
CC human islet cell antigen 1851. It was isolated from human insulinoma cDNA
CC by PCR amplification (see AAT95224-25). Other partial clones (see
CC AAT95226 and AAT95228) were similarly obtained. Subsequent 5' RACE
CC yielded a claimed sequence (see AAT95221) for islet cell antigen 1851
CC (see AAW35297). This antigen forms an immune complex with an autoantibody
CC found in patients at risk of, or predisposed to, insulin-dependent
CC diabetes mellitus (IDDM), and can be used in diagnostic and therapeutic
CC methods
XX
SQ Sequence 2328 BP; 597 A; 599 C; 595 G; 537 T; 0 U; 0 Other;
Query Match 93.7%; Score 426.4; DB 2; Length 2328;
Best Local Similarity 98.9%; Pred. No. 3.3e-110;
Matches 449; Conservative 0; Mismatches 3; Indels 2; Gaps 2;
QY 1 GAAATCAGAAGTTTAAATATGACACAAATTAATATATTTGTATATCTCACACCGGA-GNNTT 59
DB 2282 GAATCAGAAGTTTAAATATGACACAAATTAATATATTTGTATATCTCACACGGAGTTT 2223
QY 60 CTCCTTCAACATAAGGAGTTAGAAATTAACAAGTAGGCATATGCTTCTTATATTCAGATAA 119
DB 2222 CTCCTTCAACATAAGGAGTTAGAAATTAACAAGTAGGCATATGCTTCTTATATTCAGATAA 2163
QY 120 ATTCATTTTCGATTAATTAATTCAGATAGAGAGAGTAATTTTCGGAAGAAATGATA 179
DB 2162 ATTCATTTTCGATTAATTAATTCAGATAGAGAGAGTAATTTTCGGAAGAAATGATA 2104
QY 180 GCTATATTAAAGCAGATATTCTATTACATACCATCTAGACATAGCAATATTTTGGCA 239
DB 2103 GCTATATTAAAGCAGATATTCTATTACATACCATCTAGACATAGCAATATTTTGGCA 2044
QY 240 TCATTTCTGCTCCAGTAGGCGGTGTTCCCTCTCGTAGGCGCTTTGGAGAGTACCATCT 299
DB 2043 TCATTTCTGCTCCAGTAGGCGGTGTTCCCTCTCGTAGGCGCTTTGGAGAGTACCATCT 1984
QY 300 ATCTAAGATGGAGGAATGCTGTGGGAAGCGGGATAGGAGTGCGTTTTCTACGCTGAAC 359
|||||

DB 1383 ATCTAAGATGGAGGAATGCTGTGGGAAGCGGGATGGAGTGCGTTTTCTACGCTGAAC 1924
QY 360 CCCACACAGGAAATCTGCAGCCACACAGCTGCTCTGCGCGCTTCCATGTGATCATC 419
DB 1923 CCCACACAGGAAATCTGCAGCCACACAGCTGCTCTGCGCGCTTCCATGTGATCATC 1864
QY 420 CTGCTCAATGAAGTAATTCCTTATTCNGGGG 453
DB 1863 CTGCTCAATGAAGTAATTCCTTATTCNGGGG 1830
RESULT 4
ADD14680/C
ID ADD14680 standard; cDNA; 4699 BP.
XX
XX AC
XX ADD14680;
XX
XX 01-JAN-2004 (first entry)
XX Human src biomarker polynucleotide SEQ ID NO:74.
DE predictor set; protein tyrosine kinase activity modulator;
KW protein tyrosine kinase pathway; protein tyrosine kinase; cytostatic;
KW gene therapy; drug sensitivity; genetic profile; cancer; human; gene; ss.
XX Homo sapiens.
OS
XX WO2003062395-A2.
PN
XX 31-JUL-2003.
PD
XX 17-JAN-2003; 2003WO-US001981.
PF
XX 18-JAN-2002; 2002US-0350061P.
PR
XX (BRIM) BRISTOL-MYERS SQUIBB CO.
PA
XX Huang F, Fairchild CR, Lee FY, Shaw P;
PI WPI; 2003-636735/60.
XX P-PSDB; ADD14084.
DR
XX New polynucleotides and polypeptides for predicting the activity of
PT compounds that interact with protein tyrosine kinases and/or protein
PT tyrosine kinase pathways.
XX
XX Claim 2; SEQ ID NO 74; 139pp; English.
PS
XX The present invention describes a predictor set comprising a plurality of
CC polynucleotides or polypeptides whose expression pattern is predictive of
CC the response of cells to treatment with a compound that modulates protein
CC tyrosine kinase activity or members of the protein tyrosine kinase
CC pathway. Also described: (1) predicting whether a compound is capable of
CC modulating the activity of cells, comprising obtaining a sample of cells,
CC determining whether the cells express a plurality of markers, and
CC correlating the expression of the markers to the compound's ability to
CC modulate the activity of the cells; (2) a plurality of cell lines for
CC identifying polynucleotides and polypeptides whose expression levels
CC correlate with compound sensitivity or resistance of cells associated
CC with a disease state; and (3) identifying polynucleotides and
CC polypeptides that predict compound sensitivity or resistance of cells
CC associated with a disease state, comprising subjecting the plurality of
CC cell lines to one or more compounds, analysing the expression pattern of
CC a microarray of polynucleotides or polypeptides, and selecting
CC polynucleotides or polypeptides that predict the sensitivity or
CC resistance of cells associated with a disease state by using the
CC expression pattern of the microarray. The polynucleotides and
CC polypeptides have cytostatic activities, and can be used in gene therapy.
CC The polynucleotides and polypeptides are useful in predicting the
CC activity of compounds that interact with protein tyrosine kinases and/or
CC protein tyrosine kinase pathways. These may be used in determining drug
CC sensitivity in patients to allow the development of individualized
CC genetic profiles which aid in treating diseases and disorders (e.g.


```
RESULT 6
ADP18660/c
ID ADP18660 standard; cDNA; 4719 BP.
XX
AC ADP18660;
XX
DT 26-AUG-2004 (first entry)
XX
DE Human TAT418 cDNA used to treat cancer SeqID 20.
XX
KW tumour; growth inhibitory; cytotoxic; cytostatic; gene therapy; cancer;
KW human; TAT; gene; ss.
XX
OS Homo sapiens.
XX
PN WO2004045516-A2.
XX
PD 03-JUN-2004.
XX
PF 13-NOV-2003; 2003WO-US036298.
XX
PR 15-NOV-2002; 2002US-0426847P.
PR 06-DEC-2002; 2002US-0431250P.
PR 31-DEC-2002; 2002US-0437344P.
XX
PA (GETH ) GENENTECH INC.
XX
PI Ashkenazi AJ, Frantz G, Goddard A, Gonzalez L, Gurney AL;
PI Polakis P, Polson A, Wood WI, Wu TD, Zhang Z;
XX
DR WPI; 2004-420515/39.
DR P-PSDB; ADP18681.
XX
PT New antibody that binds to tumor-associated antigenic target polypeptide
PT (TAT), useful in preparing a composition for diagnosing or treating
PT tumor.
XX
PS Claim 1; SEQ ID NO 20; 183pp; English.
XX
CC This invention relates to novel isolated antibodies that are useful for
CC the diagnosis and treatment of a tumour. Specifically, it refers to DNA
CC that is detectably labelled and conjugated to a growth inhibitory or
CC cytotoxic agent and comprises toxins, antibiotics, radioactive isotopes
CC or nucleolytic enzymes. The present invention refers to the toxin
CC maytansinoid or calicheamicin, where the host cell is one of chinese
CC hamster ovary (CHO), yeast or Escherichia coli cell. Furthermore, these
CC pharmaceutical compositions are cytostatic and can be used for gene
CC therapy purposes to treat various cancers. This polynucleotide sequence
CC is a human TAT cDNA sequence of the invention.
XX
SQ Sequence 4719 BP; 1057 A; 1404 C; 1340 G; 918 T; 0 U; 0 Other;

Query Match 91.1%; Score 414.6; DB 12; Length 4719;
Best Local Similarity 98.2%; Pred. No. 9.6e-107;
Matches 439; Conservative 0; Mismatches 6; Indels 2; Gaps 2;

QY 8 GAAAGTTTAAATGACACAAATTAATATATTTGTATATCTCACACCGGA-GNITCTCTTCA 66
DB |||||||
DB 4710 GAAAGTTTAAATGACACAAATTAATATATTTGTATATCTCACACCGAGGTTCTCTTCA 4651
QY 67 AACATAAGGAGTTGAAATTAACAGTAGGCATATCTTCTATATTCAGATAAATTCATT 126
DB |||||||
DB 4650 AACATAAGTAGTTGAAATTAACAGTAGGCATATCTTCTATATTCAGATAAATTCATT 4591
QY 127 TCGATTAAATTAATTCAGATAGAGAGAGTAATTTTCGAAAAGAAATGATAGCTATAT 186
DB |||||||
DB 4590 TCGATTAAATTAATTCAGATAGAGAGAGTAATTTTCGAAAAGAAATGATAGCTATAT 4532
QY 187 TAAAGCAGATATTCATTACATACCATGATAGAGACATAGCAATATTTTGGCATCATCT 246
DB |||||||
DB 4531 TAAAGCAGATATTCATTACATACCATGATAGAGACATAGCAATATTTTGGCATCATCT 4472
```

```
QY 247 GTCGCTCAGTAGGCGGCTTCCCTCTGCTAGGCGCTTGGAGAGTACCATCTATCTAAG 306
DB |||||||
DB 4471 GTCGCTCAGTAGGCGGCTTCCCTCTGCTAGGCGCTTGGAGAGTACCATCTATCTAAG 4412
QY 307 ATGAGGAATCTGTGGGAAGCGGGATGAGGTGCGTTTCTACGCTGAACCCACAC 366
DB |||||||
DB 4411 ATGAGGAATCTGTGGGAAGCGGGATGAGGTGCGTTTCTACGCTGAACCCACAC 4352
QY 367 AGGAAATCTGCAGCCACACAGCTGCTCTCGCGCGCTTCCATGTGATCATCTCTGGTCA 426
DB |||||||
DB 4351 AGGAAATCTGCAGCCACACAGCTGCTCTCGCGCGCTTCCATGTGATCATCTCTGGTCA 4292
QY 427 ATGAGGAATCTGCTATTTTCNGGGG 453
DB |||||||
DB 4291 ATGAGGAATCTGCTATTTTCNGGGG 4265

RESULT 7
ABV95893/c
ID ABV95893 standard; cDNA; 282 BP.
XX
AC ABV95893;
XX
DT 14-JAN-2003 (first entry)
XX
DE Human pancreatic cancer expressed cDNA SEQ ID NO 1301.
XX
KW Human; pancreas; cancer; gene therapy; vaccine; immunostimulant;
KW cytostatic; tumour; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200260317-A2.
XX
PD 08-AUG-2002.
XX
PF 30-JAN-2002; 2002WO-US002781.
XX
PR 30-JAN-2001; 2001US-0265305P.
PR 31-JAN-2001; 2001US-0265682P.
PR 09-FEB-2001; 2001US-0267568P.
PR 21-MAR-2001; 2001US-0278651P.
PR 28-APR-2001; 2001US-0287112P.
PR 16-MAY-2001; 2001US-0291631P.
PR 12-JUL-2001; 2001US-0305484P.
PR 20-AUG-2001; 2001US-0313999P.
PR 27-NOV-2001; 2001US-0333626P.
XX
PA (CORI-) CORIXA CORP.
XX
PI Benson DR, Kalos MD, Lodes MJ, Persing DH, Hepler WT, Jiang Y;
XX
DR WPI; 2002-627435/67.
XX
PT New isolated polynucleotide and pancreatic tumor polypeptides, useful for
PT diagnosing, preventing and/or treating cancer, particularly pancreatic
PT cancer.
XX
PS Claim 1; SEQ ID NO 1301; 300pp + Sequence Listing; English.
XX
CC The invention relates to an isolated polynucleotide (I) comprising: (a)
CC any of a group of over 4000 nucleotide sequences (ABV94628-ABV99145); (b)
CC complements of (a); (c) sequences consisting of at least 20 contiguous
CC residues of (a); (d) sequences that hybridize to (a), under moderately
CC stringent conditions; (e) sequences having at least 75% or 90% identity
CC to (a); or (f) degenerate variants of (a). Polypeptides (ABP68536-
CC ABP68637) encoded by (I) and oligonucleotide can be used to detect cancer
CC in a patient and compositions comprising polypeptides, polynucleotides,
CC antibodies, fusion proteins, T cell populations and antigen presenting
CC cells expressing the polypeptide are useful in treating pancreatic cancer
CC and stimulating an immune response. The polynucleotides can be used as
CC probes or primers for nucleic acid hybridisation, in the design and
CC preparation of ribozyme molecules for inhibiting expression of the tumour
```

CC	polypeptides and proteins in the tumour cells, in vaccines and for gene therapy. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences	CC	and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (III) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic coding sequences of the invention. Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
SQ	Sequence 282 BP; 92 A; 49 C; 44 G; 97 T; 0 U; 0 Other;	SQ	Sequence 4856 BP; 1082 A; 1454 C; 1374 G; 946 T; 0 U; 0 Other;
Query Match	56.0%; Score 254.8; DB 6; Length 282;	Query Match	49.5%; Score 225.2; DB 5; Length 4856;
Best Local Similarity	98.2%; Pred. No. 5.8e-62;	Best Local Similarity	82.4%; Pred. No. 4.4e-53;
Matches 278; Conservative	0; Mismatches 3; Indels 2; Gaps 2;	Matches 394; Conservative	0; Mismatches 60; Indels 24; Gaps 11;
QY	11 GTTTAATATGACACAATTAATAATATTTGTATATCTCACACCGGA-GNTTCTCTCAAAC 69	QY	1 GAAATCAGAAGTTTAAATATATGACACAATTAATAATATTTGTATATCTCACACCGG-AGNTT 59
Db	282 GTTTAATATGACACAATTAATAATATTTGTATATCTCACACCGGAGTTTCTCTCAAAC 223	Db	4833 GAAATCAGAAGTTTAAATATATGACACAATTAATAATATTTGTATATCTCACACCGGAAGTTT 4774
QY	70 ATAAGGAGTTAGAAATTAACAAGTAGGCATATGCTTCTATATTCAGATAAAATTCATTTCG 129	QY	60 CTCCTCAACACATAAGGAGTTA-GAAATACAGTAGGATAGCATATGCTTCTATATTCAGATA 118
Db	222 ATAAGTAGTGTAGAAATTAACAAGTAGGCATATGCTTCTATATTCAGATAAAATTCATTTCG 163	Db	4773 CTCCTCAACACATAAGGAGTTTAAATATTAACAAGTAGGATATGCTTCTATATTCAGATA 4714
QY	130 ATTAATTAATTCAGATAGAGAGAGTAATTTTCGAAAAGAAATGATAGCTATATTTAA 189	QY	119 AATTCATTTTCGATTAATTAATTTCCAGATAGAGAGAGTAATTTTCGGAAGAAAG-AAATGA 177
Db	162 ATTAATTAATTCAGATAGAGAGAGTAATTTTCGAAAAGAAATGATAGCTATATTTAA 104	Db	4713 AATTCATTTTCGATTAATTAATTTCCAGATAGAGAGAGTAATTTTCGGAAGAAAGCAAATGA 4654
QY	190 AGCAGATATTCATTACAATACCATGATAGAGACATAAGCAATATTTTGGCATCATTTCTGTC 249	QY	178 TAGCTATATTAAG--CAGATATTCATTACAATACCATGATAGAGACATAAG--CAATATT 233
Db	103 AGCAGATATTCATTACAATACCATGATAGAGACATAAGCAATATTTTGGCATCATTTCTGTC 44	Db	4653 TAGCTATATTAAGCGCAGGATATTCATTACAATACCATGATAGAGACATAAGCGCAATATT 4594
QY	250 CGCTCAGTAGGCGGTGCTCCCTCTGCTAGGCGCTTTGGAGAGT 292	QY	234 TTGGGATCATCTCTGTCGGCT----CAGTAGGCGGTGTCCTCTCTGGTAGGGCC--TTTGG 286
Db	43 CGCTCAGTAGGCGGTGCTCCCTCTGCTAGGCGCTTTGGAGAGT 1	Db	4593 TGGGATCATCTCTGTCGGCTTTCAGTAGGCGGTGTTTCCCTCTGGGTAGGGGCTTTTGG 4534
RESULT 8		QY	287 GAGAGTACCATCTATCTAAGATGGAGGAATGCTGTGGGAAGCGCGGATGGAGGTG---C 343
AAS87633/c		Db	4533 GAGAGTACCATCTATCTAAGATGGAGGAATGCTGTGGGAAGCGCGGATGGAGGTGCGC 4474
ID AAS87633 standard; cDNA; 4856 BP.		QY	344 GTTTTCTACGCTGAACCCACACACA--GGAATCTGACGCCACACAGC-TGCCTCTGCGC 400
AC AAS87633;		Db	4473 GTTTTCTACGCTGAACCCACACAGGAATCTGCGAGCCCAACACAGCTTGCTCTGGGG 4414
DT 13-FEB-2002 (first entry)		QY	401 CGCCTTCCATGTGATCATCTCT---GGTCAATGAAGTGAATTTGCTATTTTCNGGGG 454
DE DNA encoding novel human diagnostic protein #23437.		Db	4413 CGCCTTTCATGGTGATCATCTCTGGGTCAATGAAGTGAATTTGCTATTTTCTGGGG 4356
XX Human; chromosome mapping; gene mapping; gene therapy; forensic; food supplement; medical imaging; diagnostic; genetic disorder; ss.		RESULT 9	
OS Homo sapiens.		ADF86738	
XX WO200175067-A2.		ID ADF86738 standard; DNA; 121 BP.	
XX 11-OCT-2001.		XX	
XX 30-MAR-2001; 2001WO-US008631.		AC ADF86738;	
PR 31-MAR-2000; 2000US-00540217.		XX	
PR 23-AUG-2000; 2000US-00649167.		DT 26-FEB-2004 (first entry)	
XX (HYSE-) HYSEQ INC.		XX	
XX Drmanac RT, Liu C, Tang YT;		DE Single nucleotide polymorphism detection human gene, SEQ ID No 321.	
XX WPI; 2001-639362/73.		XX human; single nucleotide polymorphism; microarray; side effect; gene; ds.	
XX P-PSDB; ABG23446.		XX Homo sapiens.	
XX New isolated polynucleotide and encoded polypeptides, useful in		OS	
PT diagnostics, forensics, gene mapping, identification of mutations		XX JP2003235571-A.	
PT responsible for genetic disorders or other traits and to assess			
PT biodiversity.			
PS Claim 1; SEQ ID NO 23437; 103pp; English.			
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)			
CC sequences. (I) is useful as hybridisation probes, polymerase chain			
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,			

XX PD 26-AUG-2003.
 XX PF 12-FEB-2002; 2002JP-00034717.
 XX PR 12-FEB-2002; 2002JP-00034717.
 XX PA (KAGA-) KAGAKU GIJUTSU SHINKO JIGYODAN.
 XX DR WPI; 2003-820454/77.
 XX PT Novel polynucleotide useful for detecting single nucleotide polymorphisms in human gene.
 XX PS Claim 1; SEQ ID NO 321; 704pp; Japanese.
 XX CC The invention relates to a novel polynucleotide isolated and purified from a human gene having any one of 935 fully defined sequences as given in specification, or a sequence having a base substitution. The invention further relates to: an oligonucleotide containing single nucleotide polymorphisms; a PCR primer set chosen from the combination of two DNA fragments from any one of 1220 fully defined sequences as given in specification; a labelling probe containing the SNP containing oligo; and a microarray equipped with the SNP containing oligo. The isolated human gene of the invention is useful for detecting the single nucleotide polymorphisms in human gene. The isolated human gene is also useful for diagnosis of disease and determination of side effect to a medical agent. The isolated human gene is also effective in detecting single nucleotide polymorphisms in a human gene. This polynucleotide sequence represents one of the 935 isolated polynucleotides from a human gene of the invention.

XX SQ Sequence 121 BP; 27 A; 33 C; 34 G; 26 T; 0 U; 1 Other;
 Query Match 26.5%; Score 120.6; DB 10; Length 121;
 Best Local Similarity 99.2%; Pred. No. 4.4e-24;
 Matches 120; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 292 TACCATCTATCTAAGATGGAGGAATGCTGTGGAGGGCGGATGGAGTGGTGGTCTTA 351
 DB 1 TACCATCTATCTAAGATGGAGGAATGCTGTGGAGGGCGGATGGAGTGGTGGTCTTA 60

QY 352 CGTGAACCCACACAGAGAAATCTGCAGCCACACAGCTGCCTCTGCGCGGCTTCCATG 411
 DB 61 YGCTGAACCCACACAGAGAAATCTGCAGCCACACAGCTGCCTCTGCGCGGCTTCCATG 120

QY 412 T 412
 DB 121 T 121

RESULT 10
 ADF86737
 ID ADF86737 standard; DNA; 121 BP.
 AC ADF86737;
 DT 26-FEB-2004 (first entry)
 DE Single nucleotide polymorphism detection human gene, SEQ ID No 320.
 KW human; single nucleotide polymorphism; microarray; side effect; gene; ds.
 XX Homo sapiens.
 OS JP2003235571-A.
 PN 26-AUG-2003.
 XX PT Novel polynucleotide useful for detecting single nucleotide polymorphisms in human gene.
 XX PS Claim 1; SEQ ID NO 318; 704pp; Japanese.
 XX CC The invention relates to a novel polynucleotide isolated and purified from a human gene having any one of 935 fully defined sequences as given

PA (KAGA-) KAGAKU GIJUTSU SHINKO JIGYODAN.
 XX WPI; 2003-820454/77.
 XX PT Novel polynucleotide useful for detecting single nucleotide polymorphisms in human gene.
 XX PS Claim 1; SEQ ID NO 320; 704pp; Japanese.
 XX CC The invention relates to a novel polynucleotide isolated and purified from a human gene having any one of 935 fully defined sequences as given in specification, or a sequence having a base substitution. The invention further relates to: an oligonucleotide containing single nucleotide polymorphisms; a PCR primer set chosen from the combination of two DNA fragments from any one of 1220 fully defined sequences as given in specification; a labelling probe containing the SNP containing oligo; and a microarray equipped with the SNP containing oligo. The isolated human gene of the invention is useful for detecting the single nucleotide polymorphisms in human gene. The isolated human gene is also useful for diagnosis of disease and determination of side effect to a medical agent. The isolated human gene is also effective in detecting single nucleotide polymorphisms in a human gene. This polynucleotide sequence represents one of the 935 isolated polynucleotides from a human gene of the invention.

XX SQ Sequence 121 BP; 25 A; 34 C; 30 G; 31 T; 0 U; 1 Other;
 Query Match 25.2%; Score 114.6; DB 10; Length 121;
 Best Local Similarity 98.3%; Pred. No. 2.2e-22;
 Matches 114; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 338 AGTGCGGTTTCTAGCGTGAACCCACACAGAGAAATCTGCAGCCACACAGCTGCCTCTG 397
 DB 1 AGTGCGGTTTCTAGCGTGAACCCACACAGAGAAATCTGCAGCCACACAGCTGCCTCTG 60

QY 398 CGCGCGCTTCCATGTGATCATCTGGTCAATGAATGTCTATTCNGGGG 453
 DB 61 YGCGCGCTTCCATGTGATCATCTGGTCAATGAATGTCTATTCNGGGG 116

RESULT 11
 ADF86735
 ID ADF86735 standard; DNA; 121 BP.
 AC ADF86735;
 DT 26-FEB-2004 (first entry)
 DE Single nucleotide polymorphism detection human gene, SEQ ID No 318.
 KW human; single nucleotide polymorphism; microarray; side effect; gene; ds.
 XX Homo sapiens.
 OS JP2003235571-A.
 PN 26-AUG-2003.
 XX PT Novel polynucleotide useful for detecting single nucleotide polymorphisms in human gene.
 XX PS Claim 1; SEQ ID NO 318; 704pp; Japanese.
 XX CC The invention relates to a novel polynucleotide isolated and purified from a human gene having any one of 935 fully defined sequences as given

CC in specification, or a sequence having a base substitution. The invention
CC further relates to: an oligonucleotide containing single nucleotide
CC polymorphisms; a PCR primer set chosen from the combination of two DNA
CC fragments from any one of 1220 fully defined sequences as given in
CC specification; a labelling probe containing the SNP containing oligo; and
CC a microarray equipped with the SNP containing oligo. The isolated human
CC gene of the invention is useful for detecting the single nucleotide
CC polymorphisms in human gene. The isolated human gene is also useful for
CC diagnosis of disease and determination of side effect to a medical agent.
CC The isolated human gene is also effective in detecting single nucleotide
CC polymorphisms in a human gene. This polynucleotide sequence represents
CC one of the 935 isolated polynucleotides from a human gene of the
CC invention.
XX
SQ Sequence 121 BP; 50 A; 14 C; 17 G; 39 T; 0 U; 1 Other;

Query Match 24.1%; Score 109.6; DB 10; Length 121;
Best Local Similarity 98.4%; Pred. No. 5.7e-21;
Matches 120; Conservative 1; Mismatches 0; Indels 1; Gaps 1;
QY 89 AAGTAGGCATATGCTTCCTATATTCAGATAAAATTCATTTCGATTAAATTAATTCAGATA 148
DB 1 AAGTAGGCATATGCTTCCTATATTCAGATAAAATTCATTTCGATTAAATTAATTCAGATA 60
QY 149 GAGAGAAGTAATTTTCGAAAAAGAAATGATAGCTATATTAAGCAGATATTCATTACAAT 208
DB 61 RAGAGAAGTAATTTT-GGAAAGAAATGATAGCTATATTAAGCAGATATTCATTACAAT 119
QY 209 AC 210
DB 120 AC 121

RESULT 12
ADP86734
ID ADF86734 standard; DNA; 121 BP.
XX AC
XX ADF86734;
XX
DT 26-FEB-2004 (first entry)
XX
DE Single nucleotide polymorphism detection human gene, SEQ ID No 317.
XX
KW human; single nucleotide polymorphism; microarray; side effect; gene; ds.
XX
OS Homo sapiens.
XX
PN JP2003235571-A.
XX
PD 26-AUG-2003.
XX
PF 12-FEB-2002; 2002JP-00034717.
XX
PR 12-FEB-2002; 2002JP-00034717.
XX
PA (KAGA-) KAGAKU GIJUTSU SHINKO JIGYODAN.
XX
PS WPI; 2003-820454/77.
XX
DR
XX

CC The invention relates to a novel polynucleotide isolated and purified
CC from a human gene having any one of 935 fully defined sequences as given
CC in specification, or a sequence having a base substitution. The invention
CC further relates to: an oligonucleotide containing single nucleotide
CC polymorphisms; a PCR primer set chosen from the combination of two DNA
CC fragments from any one of 1220 fully defined sequences as given in
CC specification; a labelling probe containing the SNP containing oligo; and
CC a microarray equipped with the SNP containing oligo. The isolated human
CC gene of the invention is useful for detecting the single nucleotide

CC polymorphisms in human gene. The isolated human gene is also useful for
CC diagnosis of disease and determination of side effect to a medical agent.
CC The isolated human gene is also effective in detecting single nucleotide
CC polymorphisms in a human gene. This polynucleotide sequence represents
CC one of the 935 isolated polynucleotides from a human gene of the
CC invention.
XX
SQ Sequence 121 BP; 44 A; 19 C; 15 G; 42 T; 0 U; 1 Other;

Query Match 23.6%; Score 107.6; DB 10; Length 121;
Best Local Similarity 97.5%; Pred. No. 2.1e-20;
Matches 118; Conservative 1; Mismatches 1; Indels 1; Gaps 1;
QY 16 ATATGACACAATTAATATATTTGTATATCTCACACCGGA-GNTTCTCTTCAACATAAG 74
DB 1 ATATGACACAATTAATATATTTGTATATCTCACACCGGAGTTCTCTTCAACATAAG 60
QY 75 GAGTTAGAAATTAAGTAGGCATATGCTTCCTATATTCAGATAAATTCATTTCGATTAA 134
DB 61 KAGTTAGAAATTAAGTAGGCATATGCTTCCTATATTCAGATAAATTCATTTCGATTAA 120
QY 135 T 135
DB 121 T 121

RESULT 13
ADP86736
ID ADF86736 standard; DNA; 121 BP.
XX AC
XX ADF86736;
XX
DT 26-FEB-2004 (first entry)
XX
DE Single nucleotide polymorphism detection human gene, SEQ ID No 319.
XX
KW human; single nucleotide polymorphism; microarray; side effect; gene; ds.
XX
OS Homo sapiens.
XX
PN JP2003235571-A.
XX
PD 26-AUG-2003.
XX
PF 12-FEB-2002; 2002JP-00034717.
XX
PR 12-FEB-2002; 2002JP-00034717.
XX
PA (KAGA-) KAGAKU GIJUTSU SHINKO JIGYODAN.
XX
PS WPI; 2003-820454/77.
XX
DR
XX
PT Novel polynucleotide useful for detecting single nucleotide polymorphisms
PT in human gene.
XX
PS Claim 1; SEQ ID NO 319; 704pp; Japanese.

CC The invention relates to a novel polynucleotide isolated and purified
CC from a human gene having any one of 935 fully defined sequences as given
CC in specification, or a sequence having a base substitution. The invention
CC further relates to: an oligonucleotide containing single nucleotide
CC polymorphisms; a PCR primer set chosen from the combination of two DNA
CC fragments from any one of 1220 fully defined sequences as given in
CC specification; a labelling probe containing the SNP containing oligo; and
CC a microarray equipped with the SNP containing oligo. The isolated human
CC gene of the invention is useful for detecting the single nucleotide
CC polymorphisms in human gene. The isolated human gene is also useful for
CC diagnosis of disease and determination of side effect to a medical agent.
CC The isolated human gene is also effective in detecting single nucleotide
CC polymorphisms in a human gene. This polynucleotide sequence represents
CC one of the 935 isolated polynucleotides from a human gene of the
CC invention.

SQ Sequence 121 BP; 26 A; 33 C; 32 G; 29 T; 0 U; 1 Other;
Query Match 20.4%; Score 92.6; DB 10; Length 121;
Best Local Similarity 97.9%; Pred. No. 3.7e-16;
Matches 92; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 360 CCCACACAGGAATCTGACGCCACACAGCTGCTCTGGCGCGCTCCATGTGATCATC 419
DB 1 CCCACACAGGAATCTGACGCCACACAGCTGCTCTGGCGCGCTCCATGTGATCATC 60
QY 420 CTGGTCAATGAAGTGAATGTCTATTTCNCGGG 453
DB 61 YGGTCAATGAAGTGAATGTCTATTTCNCGGG 94
RESULT 14
AAH35240/c
ID AAH35240 standard; cDNA; 455 BP.
XX
AC AAH35240;
XX
DT 03-SEP-2001 (first entry)
XX
DE Human colon cancer antigen encoding cDNA SEQ ID NO:2322.
XX
DE Human colon cancer; colon cancer antigen; diagnosis; detection;
KW colorectal carcinoma; ss.
XX
OS Homo sapiens.
XX
PN WO200122920-A2.
XX
PD 05-APR-2001.
XX
PF 28-SEP-2000; 2000WO-US026524.
XX
PR 29-SEP-1999; 99US-0157137P.
PR 03-NOV-1999; 99US-0163280P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Ruben SM, Barash SC, Birse CE, Rosen CA;
XX
DR WPI; 2001-235357/24.
DR P-PSDB; AAG75835.
PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
PT useful for preventing, diagnosing and/or treating colorectal cancers.
XX
PS Claim 1; Page 3851; 9803pp; English.
XX
CC AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
CC cancer-associated nucleic acid molecules (N) and proteins (P), where the
CC proteins are collectively known as colon cancer antigens. The colon
CC cancer antigens have cytostatic activity and can be used in gene therapy
CC and vaccine production. N and P may be used in the prevention, diagnosis
CC and treatment of diseases associated with inappropriate P expression. For
CC example, N and P may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of P by expressing inactive proteins or to
CC supplement the patient's own production of P. Additionally, N may be used
CC to produce the colon cancer-associated P, by inserting the nucleic acids
CC into a host cell and culturing the cell to express the proteins. N and P
CC can be used in the prevention, diagnosis and treatment of colorectal
CC carcinomas and cancers. AAH37196 to AAH37204 and AAG77789 represent
CC sequences used in the exemplification of the present invention. N.B.
CC Pages 666 to 682 and page 7053 of the sequence listing were missing at
CC time of publication, meaning no sequences are present for SEQ ID NO:1027
CC to 1052, 7921 and 7922
XX
SQ Sequence 455 BP; 110 A; 119 C; 105 G; 106 T; 0 U; 15 Other;
Query Match 14.2%; Score 64.4; DB 4; Length 455;

Best Local Similarity 90.5%; Pred. No. 6.2e-08;
Matches 76; Conservative 0; Mismatches 7; Indels 1; Gaps 1;
QY 2 AAATCAGAGTTTAAATATGACACAATTAATTAATATTTGTATATCTCACACCGGA-GNTTC 60
DB 147 AAATCAGAGTTTAAATATGACACNCAATTAATTAATATTTGTATATCTCACNCCNANGTTTC 88
QY 61 TCTTCAACATAAGGAGTTAGAAA 84
DB 87 TTTTCAACATAAGGAGTAAGAAA 64
RESULT 15
ADQ53174/c
ID ADQ53174 standard; DNA; 538 BP.
XX
AC ADQ53174;
XX
DT 21-OCT-2004 (first entry)
XX
DE Novel canine microarray-related DNA sequence SeqID4476.
XX
DE canine microarray; drug screening; toxicity assay;
KW environmental pollutant; cellular response; gene expression profile;
KW toxic response; liver necrosis; fatty liver disease;
KW protein adduct formation; hepatitis; dog; ds.
XX
OS Canis familiaris.
XX
PN WO2004063324-A2.
XX
PD 29-JUL-2004.
XX
PF 05-MAY-2003; 2003WO-US013853.
XX
PR 03-MAY-2002; 2002US-0377240P.
XX
XX (GENE-) GENE LOGIC INC.
PA (PFIZ) PFIZER PROD INC.
XX
PI Diggans JC, Porter M, Wei T;
XX
DR WPI; 2004-561890/54.
XX
PT New isolated nucleic acid molecule, useful for drug screening and
PT toxicity assays or for assessing the impact, including toxicity, of a
PT compound, pharmaceutical agent or environmental pollutant on a cell or
PT living organism.
XX
PS Claim 1; SEQ ID NO 4476; 41pp; English.
XX
CC This invention is related to a novel isolated canine nucleic acid
CC sequences and the construction of canine microarrays containing a
CC significant portion of the canine genome. The isolated canine nucleic
CC acid sequences of the invention may be useful for drug screening and
CC toxicity assays. The invention is therefore useful for assessing the
CC impact, including toxicity, of a compound, pharmaceutical agent or
CC environmental pollutant on a cell or living organism. The methods are
CC useful for detecting genes that are up- or down-regulated in canines in a
CC disease state. The sequences are useful as diagnostic agents or markers
CC to detect a cellular response in a sample individually or as part of a
CC gene expression profile. It is also useful as a target for agents that
CC modulate gene expression or activity. The database is useful for
CC producing electronic Northern blots that allow the user to determine the cell
CC type or tissue in which a given gene is expressed and to allow
CC determination of the abundance or expression level of a given gene in a
CC particular tissue or cell. The methods are useful for determining the
CC similarity of a toxic response to one or more individual compounds. The
CC methods are useful for predicting at least one toxic response or the
CC likelihood that a compound or test agent will induce various specific
CC pathologies such as those of the liver (liver necrosis, fatty liver
CC disease, protein adduct formation or hepatitis), those of the kidney,
CC heart, brain or testes, or other pathologies associated with at least one

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 4, 2005, 23:46:20 ; Search time 102.218 Seconds
(without alignments)
7283.499 Million cell updates/sec

Title: US-08-731-499-8
Perfect score: 455
Sequence: 1 GAAATCAGAGCTTAATATG.....ATTGCTCTATTTCNGGGGGT 455

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA.*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PTUS_COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfiles.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	453	99.6	455	2	US-08-680-395-8
2	453	99.6	455	4	US-08-892-695-8
3	428	94.1	4541	4	US-09-949-016-5619
4	428	94.1	767677	4	US-09-949-016-12147
5	428	94.1	767677	4	US-09-949-016-17361
6	426.4	93.7	2328	3	US-08-811-481-34
7	426.4	93.7	2328	4	US-08-876-527-34
8	414.6	91.1	4719	4	US-09-949-016-405
9	45.2	9.9	263693	4	US-09-949-016-12386
10	45.2	9.9	263694	4	US-09-949-016-16915
11	43	9.5	601	4	US-09-949-016-35582
12	43	9.5	601	4	US-09-949-016-136427
13	43	9.5	2237	1	US-08-487-1358-1
14	43	9.5	2237	2	US-08-915-972A-1
15	43	9.5	2237	2	US-09-177-909-1
16	43	9.5	3317	4	US-09-949-016-3830
17	43	9.5	3323	4	US-09-949-016-913
18	43	9.5	89220	4	US-09-949-016-12655
19	43	9.5	89224	4	US-09-949-016-13572
20	42.8	9.4	601	4	US-09-949-016-10531
21	42.8	9.4	601	4	US-09-949-016-30532
22	42.8	9.4	601	4	US-09-949-016-37150
23	42.8	9.4	601	4	US-09-949-016-37151
24	42.8	9.4	601	4	US-09-949-016-37164
25	42.8	9.4	601	4	US-09-949-016-37165
26	42.8	9.4	601	4	US-09-949-016-145868
27	42.8	9.4	601	4	US-09-949-016-145869

C 28	42.8	9.4	601	4	US-09-949-016-146136	Sequence 146136,
C 29	42.8	9.4	601	4	US-09-949-016-146137	Sequence 146137,
C 30	42.8	9.4	601	4	US-09-949-016-146404	Sequence 146404,
C 31	42.8	9.4	601	4	US-09-949-016-146405	Sequence 146405,
C 32	42.8	9.4	205044	4	US-09-949-016-15851	Sequence 15851, A
C 33	42.8	9.4	205044	4	US-09-949-016-15852	Sequence 15852, A
C 34	42.8	9.4	205044	4	US-09-949-016-15853	Sequence 15853, A
C 35	42.8	9.4	223471	4	US-09-949-016-12387	Sequence 12387, A
C 36	42.8	9.4	223471	4	US-09-949-016-12725	Sequence 12725, A
C 37	42.8	9.4	223471	4	US-09-949-016-12724	Sequence 12724, A
C 38	42	9.2	601	4	US-09-949-016-196365	Sequence 196365,
C 39	42	9.2	601	4	US-09-949-016-196366	Sequence 196366,
C 40	42	9.2	601	4	US-09-949-016-196367	Sequence 196367,
C 41	42	9.2	451924	4	US-09-949-016-12896	Sequence 12896, A
C 42	42	9.2	451925	4	US-09-949-016-17305	Sequence 17305, A
C 43	41.8	9.2	68667	4	US-09-949-016-17017	Sequence 17017, A
C 44	41.6	9.1	76281	4	US-09-949-016-12708	Sequence 12708, A
C 45	40.6	8.9	157822	4	US-09-949-016-16723	Sequence 16723, A

ALIGNMENTS

RESULT 1
US-08-680-395-8
; Sequence 8, Application US/08680395
; Patent No. 5892010
; GENERAL INFORMATION:
; APPLICANT: Gray, Joe W.
; APPLICANT: Collins, Colin
; APPLICANT: Hwang, Soo-in
; APPLICANT: Godfrey, Tony
; APPLICANT: Kowbel, David
; APPLICANT: Rommens, Johanna
; TITLE OF INVENTION: Genes from the 20q13 Amplicon and Their
; TITLE OF INVENTION: Uses
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/680.395
; FILING DATE: 15-JUL-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Bastian, Kevin L.
; REGISTRATION NUMBER: 34,774
; REFERENCE/DOCKET NUMBER: 023070-0689000US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 455 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: -
; LOCATION: 1..455 /note= "cDNA clone 20sa7 for a homolog
; OTHER INFORMATION: of rat gene BEM-1"
; OTHER INFORMATION:
US-08-680-395-8

Query Match 99.6%; Score 453; DB 2; Length 455;
Best Local Similarity 100.0%; Pred. No. 6.3e-134;
Matches 455; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAAATCAGAAGTTTAATATGACACAATTAATATTTGTATATCTCACACCGGAGNTTC 60
DB 1 GAAATCAGAAGTTTAATATGACACAATTAATATTTGTATATCTCACACCGGAGNTTC 60

QY 61 TCTTCAAAACATAAGGAGTTAGAAATTAACAAGTAGGCATATGCTTCTATATTCAGATAAA 120
DB 61 TCTTCAAAACATAAGGAGTTAGAAATTAACAAGTAGGCATATGCTTCTATATTCAGATAAA 120

QY 121 TCCATTTCCGATTAATTAATTCAGATAGAGAGAAGTAATTTTCGGAAAAGAAATGATAG 180
DB 121 TCCATTTCCGATTAATTAATTCAGATAGAGAGAAGTAATTTTCGGAAAAGAAATGATAG 180

QY 181 CTATATTAAGCAGATATTCATTAACAATACCATGTAGACATATAAGCAATATTTTGGCAT 240
DB 181 CTATATTAAGCAGATATTCATTAACAATACCATGTAGACATATAAGCAATATTTTGGCAT 240

QY 241 CATTCGTCCGCTCAGTAGGCGGTGTTCCCTCTGGTAGGCGCTTTGGAGAGTACCATCTA 300
DB 241 CATTCGTCCGCTCAGTAGGCGGTGTTCCCTCTGGTAGGCGCTTTGGAGAGTACCATCTA 300

QY 301 TCTAAGATGGAGGAATCTGTGGAAAGGCGGAGTGGAGTGGTCTTCTACGCTGAACC 360
DB 301 TCTAAGATGGAGGAATCTGTGGAAAGGCGGAGTGGAGTGGTCTTCTACGCTGAACC 360

QY 361 CCACACAGGAATCTCAGCCACACAGCTGCCCTCTGCGCGCGCTTCCATGTGATCATCC 420
DB 361 CCACACAGGAATCTCAGCCACACAGCTGCCCTCTGCGCGCGCTTCCATGTGATCATCC 420

QY 421 TGGTCAATGAAGTGAATTCCTATTTCNGGGGT 455
DB 421 TGGTCAATGAAGTGAATTCCTATTTCNGGGGT 455

RESULT 2

US-08-892-695-8
; Sequence 8, Application US/08892695A
; Patent No. 6808878
; GENERAL INFORMATION:
; APPLICANT: Gray, Joe W
; APPLICANT: Collins, Collin
; APPLICANT: Hwang, Soo In
; APPLICANT: Godfrey, Tony
; APPLICANT: Kowel, David
; APPLICANT: Rommens, Johanna
; TITLE OF INVENTION: GENES FROM THE 20Q13 AMPLICON AND THEIR USES
; FILE REFERENCE: 2500.124US
; CURRENT APPLICATION NUMBER: US/08/892,695A
; PRIOR FILING DATE: 1997-07-15
; EARLIER APPLICATION NUMBER: 08/785,532
; EARLIER FILING DATE: 1997-01-17
; EARLIER APPLICATION NUMBER: 08/731,499
; EARLIER FILING DATE: 1996-10-16
; EARLIER APPLICATION NUMBER: 08/680,395
; EARLIER FILING DATE: 1996-07-15
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 455
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:20sa7
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (57)
; OTHER INFORMATION: n is A, C, G, T, or U
; FEATURE:
; NAME/KEY: modified_base

; LOCATION: (449)
US-08-892-695-8

Query Match 99.6%; Score 453; DB 4; Length 455;
Best Local Similarity 100.0%; Pred. No. 6.3e-134;
Matches 455; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAAATCAGAAGTTTAATATGACACAATTAATATTTGTATATCTCACACCGGAGNTTC 60
DB 1 GAAATCAGAAGTTTAATATGACACAATTAATATTTGTATATCTCACACCGGAGNTTC 60

QY 61 TCTTCAAAACATAAGGAGTTAGAAATTAACAAGTAGGCATATGCTTCTATATTCAGATAAA 120
DB 61 TCTTCAAAACATAAGGAGTTAGAAATTAACAAGTAGGCATATGCTTCTATATTCAGATAAA 120

QY 121 TCCATTTCCGATTAATTAATTCAGATAGAGAGAAGTAATTTTCGGAAAAGAAATGATAG 180
DB 121 TCCATTTCCGATTAATTAATTCAGATAGAGAGAAGTAATTTTCGGAAAAGAAATGATAG 180

QY 181 CTATATTAAGCAGATATTCATTAACAATACCATGTAGACATATAAGCAATATTTTGGCAT 240
DB 181 CTATATTAAGCAGATATTCATTAACAATACCATGTAGACATATAAGCAATATTTTGGCAT 240

QY 241 CATTCGTCCGCTCAGTAGGCGGTGTTCCCTCTGGTAGGCGCTTTGGAGAGTACCATCTA 300
DB 241 CATTCGTCCGCTCAGTAGGCGGTGTTCCCTCTGGTAGGCGCTTTGGAGAGTACCATCTA 300

QY 301 TCTAAGATGGAGGAATCTGTGGAAAGGCGGAGTGGAGTGGTCTTCTACGCTGAACC 360
DB 301 TCTAAGATGGAGGAATCTGTGGAAAGGCGGAGTGGAGTGGTCTTCTACGCTGAACC 360

QY 361 CCACACAGGAATCTCAGCCACACAGCTGCCCTCTGCGCGCGCTTCCATGTGATCATCC 420
DB 361 CCACACAGGAATCTCAGCCACACAGCTGCCCTCTGCGCGCGCTTCCATGTGATCATCC 420

QY 421 TGGTCAATGAAGTGAATTCCTATTTCNGGGGT 455
DB 421 TGGTCAATGAAGTGAATTCCTATTTCNGGGGT 455

RESULT 3

US-09-949-016-5619/c
; Sequence 5619, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5619
; LENGTH: 4541
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-5619

Query Match 94.1%; Score 428; DB 4; Length 4541;
Best Local Similarity 99.1%; Pred. No. 2.1e-125;
Matches 450; Conservative 0; Mismatches 2; Indels 2; Gaps 2;

QY 1 GAAATCAGAAGTTTAATATGACACAATTAATATTTGTATATCTCACACCGGA-GNTT 59
DB 4541 GAAATCAGAAGTTTAATATGACACAATTAATATTTGTATATCTCACACCGGAGTTT 4482

QY 60 CTCTTCAACATAGGAGTTAGAAATTACAGTAGGCGATATGCTTCTATATTCAGATAA 119
Db 4481 CTCTTCAACATAGGAGTTAGAAATTACAGTAGGCGATATGCTTCTATATTCAGATAA 4422
QY 120 ATTCATTTTCGATTAATTAATTTCCAGATAGAGAGAGTAATTTTCGGAAAGAAATGATA 179
Db 4421 ATTCATTTTCGATTAATTAATTTCCAGATAGAGAGAGTAATTTTCGGAAAGAAATGATA 4363
QY 180 GCTATATTTAAAGCAGATATTCATTAACAATACCATGTAGAGACATAAGCAATATTTTGGCA 239
Db 4362 GCTATATTTAAAGCAGATATTCATTAACAATACCATGTAGAGACATAAGCAATATTTTGGCA 4303
QY 240 TCATTTCTCGCTCAGTAGGCGGTTCCTCTGGTAGGGCCCTTTGGAGAGTACCATCT 299
Db 4302 TCATTTCTCGCTCAGTAGGCGGTTCCTCTGGTAGGGCCCTTTGGAGAGTACCATCT 4243
QY 300 ATCTAAGATGAGGAATCTGTGGAGGCGGAGTGGAGTGCGTTTCTACGCTGAAC 359
Db 4242 ATCTAAGATGAGGAATCTGTGGAGGCGGAGTGGAGTGCGTTTCTACGCTGAAC 4183
QY 360 CCCACACAGGAATCTGCAGGCCACACAGCTGCCTCTCGCGGCCCTTCCATGTGATCATC 419
Db 4182 CCCACACAGGAATCTGCAGGCCACACAGCTGCCTCTCGCGGCCCTTCCATGTGATCATC 4123
QY 420 CTGGTCAATGAAGTGAATTCCTATTTTCNGGGG 453
Db 4122 CTGGTCAATGAAGTGAATTCCTATTTCTGGG 4089

RESULT 4

US-09-949-016-12147/c
; Sequence 12147, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12147
; LENGTH: 767677
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(767677)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-12147

Query Match

Best Local Similarity 94.1%; Score 428; DB 4; Length 767677;
Matches 450; Conservative 0; Mismatches 2; Indels 2; Gaps 2;

QY 1 GAAATCAGAAGTTTAATATGACACAATTAATATATTTTGTATATCTCACCGGA-GNMT 59
Db 765677 GAAATCAGAAGTTTAATATGACACAATTAATATATTTTGTATATCTCACCGGAAGTTT 765618
QY 60 CTCTTCAACATAGGAGTTAGAAATTACAGTAGGCGATATGCTTCTATATTCAGATAA 119
Db 765617 CTCTTCAACATAGGAGTTAGAAATTACAGTAGGCGATATGCTTCTATATTCAGATAA 765558
QY 120 ATTCATTTTCGATTAATTAATTTCCAGATAGAGAGTAATTTTCGGAAAGAAATGATA 179
Db 765557 ATTCATTTTCGATTAATTAATTTCCAGATAGAGAGTAATTTTCGGAAAGAAATGATA 765499

QY 180 GCTATATTTAAAGCAGATATTCATTAACAATACCATGTAGAGACATAAGCAATATTTTGGCA 239
Db 765498 GCTATATTTAAAGCAGATATTCATTAACAATACCATGTAGAGACATAAGCAATATTTTGGCA 765439
QY 240 TCATTTCTCGCTCAGTAGGCGGTTCCTCTGGTAGGGCCCTTTGGAGAGTACCATCT 299
Db 765438 TCATTTCTCGCTCAGTAGGCGGTTCCTCTGGTAGGGCCCTTTGGAGAGTACCATCT 765379
QY 300 ATCTAAGATGAGGAATCTGTGGAGGCGGAGTGGAGTGCGTTTCTACGCTGAAC 359
Db 765378 ATCTAAGATGAGGAATCTGTGGAGGCGGAGTGGAGTGCGTTTCTACGCTGAAC 765319
QY 360 CCCACACAGGAATCTGCAGGCCACACAGCTGCCTCTCGCGGCCCTTCCATGTGATCATC 419
Db 765318 CCCACACAGGAATCTGCAGGCCACACAGCTGCCTCTCGCGGCCCTTCCATGTGATCATC 765259
QY 420 CTGGTCAATGAAGTGAATTCCTATTTTCNGGGG 453
Db 765258 CTGGTCAATGAAGTGAATTCCTATTTCTGGG 765225

RESULT 5

US-09-949-016-17361/c
; Sequence 17361, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17361
; LENGTH: 767677
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(767677)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-17361

Query Match

Best Local Similarity 94.1%; Score 428; DB 4; Length 767677;
Matches 450; Conservative 0; Mismatches 2; Indels 2; Gaps 2;

QY 1 GAAATCAGAAGTTTAATATGACACAATTAATATATTTTGTATATCTCACCGGA-GNMT 59
Db 765677 GAAATCAGAAGTTTAATATGACACAATTAATATATTTTGTATATCTCACCGGAAGTTT 765618
QY 60 CTCTTCAACATAGGAGTTAGAAATTACAGTAGGCGATATGCTTCTATATTCAGATAA 119
Db 765617 CTCTTCAACATAGGAGTTAGAAATTACAGTAGGCGATATGCTTCTATATTCAGATAA 765558
QY 120 ATTCATTTTCGATTAATTAATTTCCAGATAGAGAGTAATTTTCGGAAAGAAATGATA 179
Db 765557 ATTCATTTTCGATTAATTAATTTCCAGATAGAGAGTAATTTTCGGAAAGAAATGATA 765499
QY 180 GCTATATTTAAAGCAGATATTCATTAACAATACCATGTAGAGACATAAGCAATATTTTGGCA 239
Db 765498 GCTATATTTAAAGCAGATATTCATTAACAATACCATGTAGAGACATAAGCAATATTTTGGCA 765439
QY 240 TCATTTCTCGCTCAGTAGGCGGTTCCTCTGGTAGGGCCCTTTGGAGAGTACCATCT 299
Db 765439 TCATTTCTCGCTCAGTAGGCGGTTCCTCTGGTAGGGCCCTTTGGAGAGTACCATCT 765379

Db 765438 TCATTCTGTCGCTCAGTAGGCGGTGTTCCCTCTGCTAGGCGCTTTGGAGAGTACCATCT 765379
QY 300 ATCTAAGATGGAGGAATGCTGTGGGAAGCGGGATGGAGGTGCGTTTCTAGCTGAAC 359
Db 765378 AICTAAGATGGAGGAATGCTGTGGGAAGCGGGATGGAGGTGCGTTTCTAGCTGAAC 765319
QY 360 CCACACAGGAAATCTGCAGCCACACACAGCTGCCTCTGCGCGCCCTTCCATGTGATCATC 419
Db 765318 CCACACAGGAAATCTGCAGCCACACACAGCTGCCTCTGCGCGCCCTTCCATGTGATCATC 765259
QY 420 CTGGTCAATGAAGTGAATGTCTTATTTTNGGGG 453
Db 765258 CTGGTCAATGAAGTGAATGTCTTATTTCTGGGG 765225

RESULT 6
US-08-811-481-34/c
; Sequence 34, Application US/08811481
; Patent No. 6300093
; GENERAL INFORMATION:
; APPLICANT: Kindsvogel, Wayne
; APPLICANT: Jelinek, Laura J.
; APPLICANT: Sheppard, Paul O.
; APPLICANT: Hagopian, William A.
; APPLICANT: LaGasse, James M.
; TITLE OF INVENTION: ISLET CELL ANTIGEN 1851
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Zymogenetics, Inc.
; STREET: 1201 Eastlake Avenue East
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/811,481
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Lingenfelter, Susan
; REGISTRATION NUMBER: P-41,156
; REFERENCE/DOCKET NUMBER: 95-36
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-442-6675
; TELEFAX: 206-442-6678
; TELEX:
; INFORMATION FOR SEQ ID NO: 34:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2328 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
US-08-811-481-34

Query Match 93.7%; Score 426.4; DB 3; Length 2328;
Best Local Similarity 98.9%; Pred. No. 4.6e-125;
Matches 449; Conservative 0; Mismatches 3; Indels 2; Gaps 2;

QY 1 GAAATCAGAGTTTAAATGACACAAATTAATATATTTGTATATCTCACACCGGA-GNTT 59
Db 2282 GAAATCAGAGTTTAAATGACACAAATTAATATATTTGTATATCTCACACCGGAGTTT 2223
QY 60 CTCCTCAAAATAGGAGTTAGAAATTAACAAGTAGGCATATGCTTCTATATTCAGATAA 119

Db 2222 CTCCTCAAAATAGGAGTTAGAAATTAACAAGTAGGCATATGCTTCTATATTCAGATAA 2163
QY 120 ATTCAATTTTCGATTAAATTAATTAATTCAGATAGAGAGTAATTTTCGAAAAAGAAATGATA 179
Db 2162 ATTCAATTTTCGATTAAATTAATTCAGATAGAGAGTAATTTTCGAAAAAGAAATGATA 2104
QY 180 GCTATATTAAGCAGATATTCATTACAATPACCATGTAGACACATAAGCAATATTTTGGCA 239
Db 2103 GCTATATTAAGCAGATATTCATTACAATPACCATGTAGACACATAAGCAATATTTTGGCA 2044
QY 240 TCATTCTGTCGCTCAGTAGGCGGTGTTCCCTCTCTGCTAGGCGCTTTGGAGAGTACCATCT 299
Db 2043 TCATTCTGTCGCTCAGTAGGCGGTGTTCCCTCTCTGCTAGGCGCTTTGGAGAGTACCATCT 1984
QY 300 ATCTAAGATGGAGGAATGCTGTGGGAAGCGGGATGGAGGTGCGTTTCTAGCTGAAC 359
Db 1983 ATCTAAGATGGAGGAATGCTGTGGGAAGCGGGATGGAGGTGCGTTTCTAGCTGAAC 1924
QY 360 CCACACAGGAAATCTGCAGCCACACACAGCTGCCTCTGCGCGCCCTTCCATGTGATCATC 419
Db 1923 CCACACAGGAAATCTGCAGCCACACAGCTGCCTCTGCGCGCCCTTCCATGTGATCATC 1864
QY 420 CTGGTCAATGAAGTGAATGTCTTATTTTNGGGG 453
Db 1863 CTGGTCAATGAAGTGAATGTCTTATTTCTGGGG 1830

RESULT 7
US-09-876-527-34/c
; Sequence 34, Application US/09876527
; Patent No. 6627735
; GENERAL INFORMATION:
; APPLICANT: Kindsvogel, Wayne
; APPLICANT: Jelinek, Laura J.
; APPLICANT: Sheppard, Paul O.
; APPLICANT: Hagopian, William A.
; APPLICANT: LaGasse, James M.
; TITLE OF INVENTION: ISLET CELL ANTIGEN 1851
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Zymogenetics, Inc.
; STREET: 1201 Eastlake Avenue East
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/876,527
; FILING DATE: 07-Jun-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/811,481
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Lingenfelter, Susan
; REGISTRATION NUMBER: P-41,156
; REFERENCE/DOCKET NUMBER: 95-36
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-442-6675
; TELEFAX: 206-442-6678
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 34:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2328 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cdna

SEQUENCE DESCRIPTION: SEQ ID NO: 34;
US-09-876-527-34
Query Match 93.7%; Score 426.4; DB 4; Length 2328;
Best Local Similarity 98.9%; Pred. No. 4.6e-125;
Matches 449; Conservative 0; Mismatches 3; Indels 2; Gaps 2;
Qy 1 GAAATCAGAGTTTAATATGACAAATTAATATATATTTTGTATATCTCACACCGGA-GNVT 59
Db 2282 GAAATCAGAGTTTAATATGACAAATTAATATATATTTTGTATATCTCACACCGGAGTTT 2223
Qy 60 CTCITCAACATAGAGGCTTGAATATACAAATTAATATATATTTTGTATATCTCACATAA 119
Db 2222 CTCITCAACATAGAGGCTTGAATATACAAATTAATATATATTTTGTATATCTCACATAA 2163
Qy 120 ATTCAATTCGATTAATTAATTCAGATAGAGAGTAATTTTTCGAAAAAGAAATGATA 179
Db 2162 ATTCAATTCGATTAATTAATTCAGATAGAGAGTAATTTTTCGAAAAAGAAATGATA 2104
Qy 180 GCTATATTAAGCAGATATTCATTAACATACATGTAGAGACATAAGCAATATTTTGGCA 239
Db 2103 GCTATATTAAGCAGATATTCATTAACATACATGTAGAGACATAAGCAATATTTTGGCA 2044
Qy 240 TCATTCGTCCGCTCAGTAGGCGGTTCCTCTGGTAGGGCCCTTTGGAGAGTACCATCT 299
Db 2043 TCATTCGTCCGCTCAGTAGGCGGTTCCTCTGGTAGGGCCCTTTGGAGAGTACCATCT 1984
Qy 300 ATCTAAGATGAGGAATCTCTGGGAAGGGCGGATGAGGTGGCTTTTCTACGCTGAAC 359
Db 1983 ATCTAAGATGAGGAATCTCTGGGAAGGGCGGATGAGGTGGCTTTTCTACGCTGAAC 1924
Qy 360 CCCACACAGAAATCTGAGCCCAACAGCTGCTCTGCGCGGCTTCCATGTGATCATC 419
Db 1923 CCCACACAGAAATCTGAGCCCAACAGCTGCTCTGCGCGGCTTCCATGTGATCATC 1864
Qy 420 CTGCTCAATGAAGTGAATTCCTATTTTCNGGGG 453
Db 1863 CTGCTCAATGAAGTGAATTCCTATTTTCNGGGG 1830
RESULT 8
US-09-949-016-405/c
; Sequence 405, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 405
; LENGTH: 4719
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-405
Query Match 91.1%; Score 414.6; DB 4; Length 4719;
Best Local Similarity 98.2%; Pred. No. 3.9e-121;
Matches 439; Conservative 0; Mismatches 6; Indels 2; Gaps 2;
Qy 8 GAAGTTTAATATGACAAATTAATATATTTTGTATATCTCACACCGGAGTTTCTCTTCA 66
Db 4710 GAAGTTTAATATGACAAATTAATATATTTTGTATATCTCACACCGGAGTTTCTCTTCA 4651

Qy 67 AACATAAGGAGTTAGAAATACAGTAGGAGCATATGCTTCTATATTCAGATAAAATTCATT 126
Db 4650 AACATAAGTATGTTAGAAATACAGTAGGAGCATATGCTTCTATATTCAGATAAAATTCATT 4591
Qy 127 TCGATTAAATTAATTCAGATAGAGAGAAAGTAATTTTCGAAAAAGAAATAGTAGCTATAT 186
Db 4590 TCGATTAAATTAATTCAGATAGAGAGAAAGTAATTTTCGAAAAAGAAATAGTAGCTATAT 4532
Qy 187 TAAAGCAGATATTCATTAACATACATGTAGAGACATAAGCAATATTTTGGCATCAATCT 246
Db 4531 TAAAGCAGATATTCATTAACATACATGTAGAGACATAAGCAATATTTTGGCATCAATCT 4472
Qy 247 GTCGCTCAGTAGGCGGTTCCTCTGGTAGGGCCCTTTGGAGAGTACCATCTATCTAAG 306
Db 4471 GTCGCTCAGTAGGCGGTTCCTCTGGTAGGGCCCTTTGGAGAGTACCATCTATCTAAG 4412
Qy 307 ATGAGGAAATGCTCTGGGAAGGGCGGATGAGGTGGCTTTTCTACGCTGAACCCACAC 366
Db 4411 ATGAGGAAATGCTCTGGGAAGGGCGGATGAGGTGGCTTTTCTACGCTGAACCCACAC 4352
Qy 367 AGAAATCTGAGCCCAACAGCTGCTCTGCGCGGCTTCCATGTGATCATCTCTGTGTC 426
Db 4351 AGAAATCTGAGCCCAACAGCTGCTCTGCGCGGCTTCCATGTGATCATCTCTGTGTC 4292
Qy 427 ATGAGTGAATTCCTATTTTCNGGGG 453
Db 4291 ATGAGTGAATTCCTATTTTCNGGGG 4265
RESULT 9
US-09-949-016-12386
; Sequence 12386, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12386
; LENGTH: 263693
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-12386
Query Match 9.9%; Score 45.2; DB 4; Length 263693;
Best Local Similarity 50.7%; Pred. No. 0.013;
Matches 107; Conservative 0; Mismatches 104; Indels 0; Gaps 0;
Qy 10 AGTTTAATATGACAAATTAATATATTTTGTATATCTCACACCGGAGTTTCTCTCAAC 69
Db 251771 AATTATATATGATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAAT 251830
Qy 70 ATAAGGAGTTAGAAATACAGTAGGAGCATATGCTTCTATATTCAGATAAAATTCATTTCG 129
Db 251831 ATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAAT 251890
Qy 130 ATTAATTTAAATTCAGATAGAGAGAAAGTAATTTTCGAAAAAGAAATAGTAGCTATATA 189
Db 251891 ATTAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAA 251950
Qy 190 AGCAGATATTCATTAACATACCATGTAGAGA 220
Db 251951 AATATAATTTGATTACATTAATTAATATA 251981

```
RESULT 10
US-09-949-016-16915
; Sequence 16915, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16915
; LENGTH: 263694
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-16915

Query Match
Best Local Similarity 9.9%; Score 45.2; DB 4; Length 263694;
Matches 107; Conservative 0; Mismatches 104; Indels 0; Gaps 0;

QY 10 AGTTTAATATGACAAATTAATATATATTTGTTATATCTCACACGGAGNTTCTCTTCAAAAC 69
DB 251771 AATTATATGATTAAATTAATTAATATATATATATTAATTAATTAATTAATATATAT 251830

QY 70 ATAAGAGTTAGAAATCAAGTAGGCATATGCTTCTTATATTCAGATAAATTCATTTCCG 129
DB 251831 ATTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 251890

QY 130 ATTAATTAATCCAGATGAGAGAGTAATTTTCGGAAAGAAATGATAGCTATATATTA 189
DB 251891 ATATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 251950

QY 190 AGCAGATATTCATTACAAATACCATGTGAGAGA 220
DB 251951 AATATAATTTGATTACATTAAATTAATATATA 251981

RESULT 11
US-09-949-016-35582/c
; Sequence 35582, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 35582
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-35582

Query Match
Best Local Similarity 9.5%; Score 43; DB 4; Length 601;
Matches 79; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

QY 24 CAATTAATATATTTGTTATATCTCACACGGAGNTTCTCTTCAAAACATAGAGTTAGAA 83
DB 276 CAATTTTATGTTATTTGAATATCAGCAAAATTTCCATAATATATCAATTAATTTGTA 217

QY 84 ATTACAAGTAGGCATATGCTTCTTATATTCAGATAAATTCATTTGATTAATAATTC 143
DB 216 ACCACATCCAGTGTCAATGCTTACTCTTAGAGTTTCAGATGAATTTTAAATAAATA 157

QY 144 AGATAGAGAGAGTAATTTT 163
DB 156 AACTCCATAGTACTAATTTT 137

RESULT 12
US-09-949-016-136427/c
; Sequence 136427, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 136427
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-136427

Query Match
Best Local Similarity 9.5%; Score 43; DB 4; Length 601;
Matches 79; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

QY 24 CAATTAATATATTTGTTATATCTCACACGGAGNTTCTCTTCAAAACATAGAGTTAGAA 83
DB 276 CAATTTTATGTTATTTGAATATCAGCAAAATTTCCATAATATATCAATTAATTTGTA 217

QY 84 ATTACAAGTAGGCATATGCTTCTTATATTCAGATAAATTCATTTGATTAATAATTC 143
DB 216 ACCACATCCAGTGTCAATGCTTACTCTTAGAGTTTCAGATGAATTTTAAATAAATA 157

QY 144 AGATAGAGAGAGTAATTTT 163
DB 156 AACTCCATAGTACTAATTTT 137

RESULT 13
US-08-487-135B-1
; Sequence 1, Application US/08487135B
; Patent No. 5821122
; GENERAL INFORMATION:
; APPLICANT: Yannick Guillaux; Francine Jotereau;
; APPLICANT: Thierry Boon-Falleur; Sophie Lucas;
; APPLICANT: Vincent Brichard
; TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULES PEPTIDES
; TITLE OF INVENTION: WHICH FORM COMPLEXES WITH MHC MOLECULE HLA-A2
; TITLE OF INVENTION: AND USES THEREOF
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Felte & Lynch
```

STREET: 805 Third Avenue
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch 1.44 Mb storage diskette
COMPUTER: IBM PS/2
OPERATING SYSTEM: PC-DOS
SOFTWARE: Wordperfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/487,135B
FILING DATE: 07-Jun-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/487,135
FILING DATE: 07 June 1995
ATTORNEY/AGENT INFORMATION:
NAME: Hanson, No. 5821122man D
REGISTRATION NUMBER: 30,946
REFERENCE/DOCKET NUMBER: LUD 5388
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 688-9200
TELEFAX: (212) 838-3884
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2237
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-487-135B-1

Query Match 9.5%; Score 43; DB 1; Length 2237;
Best Local Similarity 56.4%; Pred. No. 0.0045;
Matches 79; Conservative 0; Mismatches 61; Indels 0; Gaps 0;
QY 24 CAATTAATATATTGTATATCTCACACGGAGNTTCTTCAACATAAGAGTTAGAA 83
DB 1388 CAATTTTATGTATTTGAATATCAGCAAAATTCAGATAATTTCCATAATATATCAATTAATTTGTA 1447
QY 84 ATTACAAGTAGCATATGCTTCTATATTCAGATAAATTCATTCGATTAATTAATTC 143
DB 1448 ACCACATCCAGTGTCTACTCTTAGAGTTAGATGAATCTTAAATTAATAAAAAA 1507
QY 144 AGATAGAGAGAGTAATTTT 163
DB 1508 AAGTCCATAGTACTAATTTT 1527

RESULT 14
US-08-915-972A-1
; Sequence 1, Application US/08915972A
; Patent No. 5986145
; GENERAL INFORMATION:
; APPLICANT: Yannick Gulloux; Francine Jotereau;
; APPLICANT: Thierry Boon-Falleur; Sophie Lucas;
; APPLICANT: Vincent Brichard
; TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULE
; TITLE OF INVENTION: COMPLEXES WITH MHC MOLECULE HLA-A2 AND USES THEREOF
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fulbright & Jaworski
; STREET: 805 Third Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch 1.44 Mb storage diskette
COMPUTER: IBM PS/2
OPERATING SYSTEM: PC-DOS
SOFTWARE: Wordperfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/487,135B
FILING DATE: 07 June 1995
ATTORNEY/AGENT INFORMATION:
NAME: Hanson, No. 5821122man D
REGISTRATION NUMBER: 30,946

APPLICATION NUMBER: US/08/915,972A
FILING DATE: August 21, 1997
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/487,135
FILING DATE: 07 June 1995
ATTORNEY/AGENT INFORMATION:
NAME: Hanson, No. 5886145man D
REGISTRATION NUMBER: 30,946
REFERENCE/DOCKET NUMBER: LUD 5388 - JEL/NDH
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 688-9200
TELEFAX: (212) 838-3884
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2237
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-915-972A-1

Query Match 9.5%; Score 43; DB 2; Length 2237;
Best Local Similarity 56.4%; Pred. No. 0.0045;
Matches 79; Conservative 0; Mismatches 61; Indels 0; Gaps 0;
QY 24 CAATTAATATATTGTATATCTCACACGGAGNTTCTTCAACATAAGAGTTAGAA 83
DB 1388 CAATTTTATGTATTTGAATATCAGCAAAATTCAGATAATTTCCATAATATATCAATTAATTTGTA 1447
QY 84 ATTACAAGTAGCATATGCTTCTATATTCAGATAAATTCATTCGATTAATTAATTC 143
DB 1448 ACCACATCCAGTGTCTACTCTTAGAGTTAGATGAATCTTAAATTAATAAAAAA 1507
QY 144 AGATAGAGAGAGTAATTTT 163
DB 1508 AAGTCCATAGTACTAATTTT 1527

RESULT 15
US-09-177-909-1
; Sequence 1, Application US/09177909
; Patent No. 5958711
; GENERAL INFORMATION:
; APPLICANT: Yannick Gulloux; Francine Jotereau;
; APPLICANT: Thierry Boon-Falleur; Sophie Lucas;
; APPLICANT: Vincent Brichard
; TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULES PEPTIDES WHICH
; TITLE OF INVENTION: FORM COMPLEXES WITH MHC MOLECULE HLA-A2 AND USES THEREOF
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fulbright & Jaworski
; STREET: 805 Third Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch 1.44 Mb storage diskette
COMPUTER: IBM PS/2
OPERATING SYSTEM: PC-DOS
SOFTWARE: Wordperfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/177,909
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/915,972
FILING DATE: August 21, 1997
APPLICATION NUMBER: 08/487,135
FILING DATE: 07 June 1995
ATTORNEY/AGENT INFORMATION:
NAME: Hanson, No. 5958711man D
REGISTRATION NUMBER: 30,946

REFERENCE/DOCKET NUMBER: LUD 5388 - JEL/NDH
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 688-9200
TELEFAX: (212) 838-3884
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2237
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-177-909-1

Query Match 9.5%; Score 43; DB 2; Length 2237;
Best Local Similarity 56.4%; Pred. No. 0.0045;
Matches 79; Conservative 0; Mismatches 61; Indels 0; Gaps 0;
QY 24 CAATTAAATATATTTGTATATCTCACACCGGAGNTTCTCTTCAACATTAAGGAGTTAGAA 83
Db 1388 CAATTTTATGTATTTTGAATATCAGCAAAATTGAAATTTCCATAAATTATCATTAATTTGTA 1447
QY 84 ATTACAAGTAGGCATATGCTTCTATATTCAGATAAATTCATTTCGATTAAATTAATTC 143
Db 1448 ACCACATCCAGTGTCTATGCTTACTCTCTTAGAGTTTCAGATGAATCTTAAATTAATAAAA 1507
QY 144 AGATAGAGAGAAGTAATTTT 163
Db 1508 AAGTCCATAGTACTAATTTT 1527

Search completed: May 5, 2005, 07:03:25
Job time : 110.552 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 5, 2005, 03:59:12 ; Search time 625.302 Seconds
(without alignments)
4449.254 Million cell updates/sec

Title: US-08-731-499-8

Perfect score: 455

Sequence: 1 GAATCAGAACTTAATG.....ATTGTCCTATTTCNGGGGT 455

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 5654200 seqs, 3057283753 residues

Total number of hits satisfying chosen parameters: 11308400

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications NA.*

```

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq.*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq.*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq.*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq.*
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq.*
15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq.*
16: /cgn2_6/ptodata/2/pubpna/US10D_PUBCOMB.seq.*
17: /cgn2_6/ptodata/2/pubpna/US10E_PUBCOMB.seq.*
18: /cgn2_6/ptodata/2/pubpna/US10F_PUBCOMB.seq.*
19: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
20: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq.*
21: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
22: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	453	99.6	455	8	US-08-731-499-8
C 2	439	96.5	1258	15	Sequence 8, Appli
C 3	426.4	93.7	2328	16	Sequence 1845, Ap
C 4	426.4	93.7	2328	9	Sequence 34, Appl
C 5	426	93.6	4767	19	Sequence 34, Appl
C 6	254.8	56.0	282	14	Sequence 435, App
C 7	64.4	14.2	455	15	Sequence 1301, Ap
C 8	44.2	9.7	6811	15	Sequence 2332, Ap
C 9	44.2	9.7	11422	15	Sequence 93, Appl
C 10	44.2	9.7	11422	17	Sequence 132, App
C 11	44.2	9.7	37515	18	Sequence 18, Appl
					Sequence 27, Appl

12	43.2	9.5	670	13	US-10-027-632-180643	Sequence 180643,
13	43.2	9.5	670	17	US-10-027-632-180643	Sequence 180643,
14	43	9.5	2815	9	US-09-925-298-217	Sequence 217, App
15	43	9.5	2815	14	US-10-102-806-217	Sequence 217, App
C 16	43	9.5	61020	17	US-10-221-714A-513	Sequence 513, App
C 17	42.4	9.3	3673778	16	US-10-312-841-1	Sequence 1, Appli
C 18	42	9.2	1132	18	US-10-653-047-5452	Sequence 5452, Ap
C 19	42	9.2	8056	18	US-10-473-126-386	Sequence 386, App
C 20	42	9.2	13574	15	US-10-311-455-1290	Sequence 1290, Ap
C 21	41.2	9.1	6988	15	US-10-311-455-2414	Sequence 2414, Ap
C 22	41.2	9.1	3673778	16	US-10-312-841-2	Sequence 2, Appli
C 23	40.6	8.9	794	18	US-10-363-345A-29543	Sequence 29543, A
C 24	40.6	8.9	794	18	US-10-363-345A-29544	Sequence 29544, A
C 25	40.6	8.9	794	19	US-10-363-483A-29543	Sequence 29543, A
C 26	40.6	8.9	794	19	US-10-363-483A-29544	Sequence 29544, A
C 27	40.4	8.8	10048	18	US-10-433-793-45	Sequence 45, Appl
C 28	40	8.8	7589	15	US-10-240-453-263	Sequence 263, App
C 29	40	8.8	8246	15	US-10-311-455-176	Sequence 176, App
C 30	39.8	8.7	413	17	US-10-424-599-90506	Sequence 90506, A
C 31	39.8	8.7	6365	15	US-10-311-455-97	Sequence 97, Appl
C 32	39.8	8.7	14316	17	US-10-221-613-407	Sequence 407, App
C 33	39.6	8.7	8771	15	US-10-311-455-1797	Sequence 1797, Ap
C 34	39.4	8.7	6436	15	US-10-311-455-1665	Sequence 1665, Ap
C 35	39.2	8.6	5695	17	US-10-221-714A-260	Sequence 260, App
C 36	39.2	8.6	17848	14	US-10-239-676-28	Sequence 28, Appl
C 37	39.2	8.6	17848	15	US-10-240-453-38	Sequence 38, Appl
C 38	39.2	8.6	17848	17	US-10-257-166-58	Sequence 58, Appl
C 39	39	8.6	547	18	US-10-363-345A-1577	Sequence 1577, Ap
C 40	39	8.6	547	18	US-10-363-345A-1578	Sequence 1578, Ap
C 41	39	8.6	547	19	US-10-363-483A-1577	Sequence 1577, Ap
C 42	39	8.6	547	19	US-10-363-483A-1578	Sequence 1578, Ap
C 43	39	8.6	846	17	US-10-282-122A-11017	Sequence 11017, A
C 44	39	8.6	5511	15	US-10-311-455-1843	Sequence 1843, Ap
C 45	39	8.6	6106	15	US-10-311-455-1446	Sequence 1446, Ap

ALIGNMENTS

RESULT 1

```

US-08-731-499-8
; Sequence 8, Application US/08731499
; Publication No. US20030148270A1
; GENERAL INFORMATION:
; APPLICANT: GRAY, Joe W.
; APPLICANT: COLLINS, Colin
; APPLICANT: HWANG, Soo-In
; APPLICANT: GODFREY, Tony
; APPLICANT: KOWBEL, David
; APPLICANT: ROMMENS, Johanna
; TITLE OF INVENTION: GENES FROM THE 20q13 AMPLICON AND THEIR
; TITLE OF INVENTION: USES
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/731.499
; FILING DATE: 16-OCT-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION NUMBER: US 08/680,395
; FILING DATE: 15-JUL-1996
; ATTORNEY/AGENT INFORMATION:

```

```
/ NAME: Hunter, Tom
/ REGISTRATION NUMBER: 38,498
/ REFERENCE/DOCKET NUMBER: 23070-068910
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (415) 576-0200
/ TELEFAX: (415) 576-0300
/ INFORMATION FOR SEQ ID NO: 8:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 455 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: cDNA
/ FEATURE:
/ NAME/KEY: -
/ LOCATION: 1..455
/ OTHER INFORMATION: /note= "cDNA clone 20sa7 for a homolog
/ OTHER INFORMATION: of rat gene BEM-1"
US-08-731-499-8

Query Match 99.6%; Score 453; DB 8; Length 455;
Best Local Similarity 100.0%; Pred. No. 3.7e-107; Indels 0; Gaps 0;
Matches 455; Conservative 0; Mismatches 0;

QY 1 GAAATCAGAAGTTTAAATATGACACAAATTAATATATTTCTATATCTTCACACCGGAGNTTC 60
DB 1 GAAATCAGAAGTTTAAATATGACACAAATTAATATATTTCTATATCTTCACACCGGAGNTTC 60
QY 61 TCTTCAACATTAAGGAGTTAGAAATTACAAAGTAGGCATATGCTTCTATATTCAGATAAA 120
DB 61 TCTTCAACATTAAGGAGTTAGAAATTACAAAGTAGGCATATGCTTCTATATTCAGATAAA 120
QY 121 TTCAATTCGATTAAATTAATTCAGATAGAGAGTAATTTTCGAAAGAAATGATAG 180
DB 121 TTCAATTCGATTAAATTAATTCAGATAGAGAGTAATTTTCGAAAGAAATGATAG 180
QY 181 CTATATTAAGCAGATATTCATTACAAATACCATGTAGACATACCAATATTTTGGCAT 240
DB 181 CTATATTAAGCAGATATTCATTACAAATACCATGTAGACATACCAATATTTTGGCAT 240
QY 241 CATTCTGTCGGCTCAGTAGCGCGTGTCCCTCTGCTGAGGCGCTTTGGAGAGTACCATCTA 300
DB 241 CATTCTGTCGGCTCAGTAGCGCGTGTCCCTCTGCTGAGGCGCTTTGGAGAGTACCATCTA 300
QY 301 TCTAAGATGGAGGAATGCTGTGGGAAGGCGGATGGAGTGCGTTTCTACGCTGAACC 360
DB 301 TCTAAGATGGAGGAATGCTGTGGGAAGGCGGATGGAGTGCGTTTCTACGCTGAACC 360
QY 361 CCACACAGGAAATCTGCAGCCACACAGCTGCTCTGCGCGCGCTTCCATGTGATCATCC 420
DB 361 CCACACAGGAAATCTGCAGCCACACAGCTGCTCTGCGCGCGCTTCCATGTGATCATCC 420
QY 421 TGGTCAATGAAGTGAATGTCCTATTTTCNGGGGT 455
DB 421 TGGTCAATGAAGTGAATGTCCTATTTTCNGGGGT 455

RESULT 2
US-10-106-698-1845/c
/ Sequence 1845, Application US/10106698
/ Publication No. US20030109690A1
/ GENERAL INFORMATION:
/ APPLICANT: Ruben et al.
/ TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide
/ FILE REFERENCE: PA005PI
/ CURRENT APPLICATION NUMBER: US/10/106,698
/ CURRENT FILING DATE: 2002-03-27
/ PRIOR APPLICATION NUMBER: PCT/US00/26524
/ PRIOR FILING DATE: 2000-09-28
/ PRIOR APPLICATION NUMBER: US 60/157,137
/ PRIOR FILING DATE: 1999-09-29
/ PRIOR APPLICATION NUMBER: US 60/163,280
/ PRIOR FILING DATE: 1999-11-03
```

```
/ NUMBER OF SEQ ID NOS: 8564
/ SOFTWARE: PatentIn Ver. 3.0
/ SEQ ID NO 1845
/ LENGTH: 1258
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (1237)..(1237)
/ OTHER INFORMATION: n equals a,t,g, or c
US-10-106-698-1845

Query Match 96.5%; Score 439; DB 15; Length 1258;
Best Local Similarity 99.3%; Pred. No. 2.5e-103;
Matches 450; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 1 GAAATCAGAAGTTTAAATATGACACAAATTAATATATTTTGTATATCTTCACACCGGAGNTTC 60
DB 1213 GAAATCAGAAGTTTAAATATGACACAAATTAATATATTTTGTATATCTTCACACCGGAGNTTC 1154
QY 61 TCTTCAACATTAAGGAGTTAGAAATTACAAAGTAGGCATATGCTTCTATATTCAGATAAA 120
DB 1153 TCTTCAACATTAAGGAGTTAGAAATTACAAAGTAGGCATATGCTTCTATATTCAGATAAA 1094
QY 121 TTCAATTCGATTAAATTAATTCAGATAGAGAGTAATTTTCGAAAGAAATGATAG 180
DB 1093 TTCAATTCGATTAAATTAATTCAGATAGAGAGTAATTTTCGAAAGAAATGATAG 1035
QY 181 CTATATTAAGCAGATATTCATTACAAATACCATGTAGACATACCAATATTTTGGCAT 240
DB 1034 CTATATTAAGCAGATATTCATTACAAATACCATGTAGACATACCAATATTTTGGCAT 975
QY 241 CATTCTGTCGGCTCAGTAGCGCGTGTCCCTCTGCTGAGGCGCTTTGGAGAGTACCATCTA 300
DB 974 CATTCTGTCGGCTCAGTAGCGCGTGTCCCTCTGCTGAGGCGCTTTGGAGAGTACCATCTA 915
QY 301 TCTAAGATGGAGGAATGCTGTGGGAAGGCGGATGGAGTGCGTTTCTACGCTGAACC 360
DB 914 TCTAAGATGGAGGAATGCTGTGGGAAGGCGGATGGAGTGCGTTTCTACGCTGAACC 855
QY 361 CCACACAGGAAATCTGCAGCCACACAGCTGCTCTGCGCGCGCTTCCATGTGATCATCC 420
DB 854 CCACACAGGAAATCTGCAGCCACACAGCTGCTCTGCGCGCGCTTCCATGTGATCATCC 795
QY 421 TGGTCAATGAAGTGAATGTCCTATTTTCNGGGG 453
DB 794 TGGTCAATGAAGTGAATGTCCTATTTTCNGGGG 762

RESULT 3
US-09-876-527-34/c
/ Sequence 34, Application US/09876527
/ Patent No. US20020102616A1
/ GENERAL INFORMATION:
/ APPLICANT: Kindsvogel, Wayne
/ Jelinek, Laura J.
/ Sheppard, Paul O.
/ Hegopian, William A.
/ Lagasse, James M.
/ TITLE OF INVENTION: ISLET CELL ANTIGEN 1851
/ NUMBER OF SEQUENCES: 34
/ CORRESPONDENCE ADDRESS:
/ ADDRESS: ZymoGenetics, Inc.
/ STREET: 1201 Eastlake Avenue East
/ CITY: Seattle
/ STATE: WA
/ COUNTRY: USA
/ ZIP: 98102
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Diskette
/ COMPUTER: IBM Compatible
/ OPERATING SYSTEM: DOS
/ SOFTWARE: FastSeq for Windows Version 2.0
```

; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/876,527
 ; FILING DATE: 07-Jun-2001
 ; CLASSIFICATION: <Unknown>
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/811,481
 ; FILING DATE: <Unknown>
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Lingenfelter, Susan
 ; REGISTRATION NUMBER: P-41,156
 ; REFERENCE/DOCKET NUMBER: 95-36
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 206-442-6675
 ; TELEFAX: 206-442-6678
 ; TELEX: <Unknown>
 ; INFORMATION FOR SEQ ID NO: 34:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2328 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: cdna
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 34:
 ; US-09-876-527-34

Query Match 93.7%; Score 426.4; DB 9; Length 2328;
 Best Local Similarity 98.9%; Pred. No. 5.9e-100;
 Matches 449; Conservative 0; Mismatches 3; Indels 2; Gaps 2;
 Qy 1 GAAATCAGAGTTTAAATATGACACAAATTAATATATTTTGTATATCTCACACGGA-GNNT 59
 Db 2282 GAAATCAGAGTTTAAATATGACACAAATTAATATATTTTGTATATCTCACACGAGGTTT 2223
 Qy 60 CTCTTCAAACATAAGGAGTTAGAAATTAACAAGTAGGCATATGCTTCTATATTCAGATAA 119
 Db 2222 CTCTTCAAACATAAGGAGTTAGAAATTAACAAGTAGGCATATGCTTCTATATTCAGATAA 2163
 Qy 120 ATTCATTTTCGATTAAATTAATTCAGATAGAGAGAAATTAATTTTCGGAAGAAATGATA 179
 Db 2162 ATTCATTTTCGATTAAATTAATTCAGATAGAGAGAAATTAATTTTCGGAAGAAATGATA 2104
 Qy 180 GCTATATTAAGCAGATATTCATTAACAATACATGTAGACATAGCAATATTTTGGCA 239
 Db 2103 GCTATATTAAGCAGATATTCATTAACAATACATGTAGACATAGCAATATTTTGGCA 2044
 Qy 240 TCATTCTGCTCGCTCAGTAGGCGGTGTTCCCTCTGTTAGGCGCTTTGGAGAGTACCATCT 299
 Db 2043 TCATTCTGCTCGCTCAGTAGGCGGTGTTCCCTCTGTTAGGCGCTTTGGAGAGTACCATCT 1984
 Qy 300 ATCTAAGATGGAGGAATGCTGTGGAGAGGCGGGATGGAGGTGGCTTTTCTACGCTGAAC 359
 Db 1983 ATCTAAGATGGAGGAATGCTGTGGAGAGGCGGGATGGAGGTGGCTTTTCTACGCTGAAC 1924
 Qy 360 CCCACACAGGAAATCTGAGGCCACACAGCTGCTCTGCGCGCGCTTCCATGTGATCATC 419
 Db 1923 CCCACACAGGAAATCTGAGGCCACACAGCTGCTCTGCGCGCGCTTCCATGTGATCATC 1864
 Qy 420 CTGGTCAATGAAGTGAATGCTCTATTTCNGGGG 453
 Db 1863 CTGGTCAATGAAGTGAATGCTCTATTTCNGGGG 1830

RESULT 4
 US-10-124-089-34/c
 ; Sequence 34, Application US/10124089
 ; Publication No. US2003016067A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Kindsvogel, Wayne
 ; Jelinek, Laura J.
 ; Sheppard, Paul O.
 ; Hagopian, William A.
 ; LaGasse, James M.
 ; TITLE OF INVENTION: ISLET CELL ANTIGEN 1851

; NUMBER OF SEQUENCES: 34
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: ZymoGenetics, Inc.
 ; STREET: 1201 Eastlake Avenue East
 ; CITY: Seattle
 ; STATE: WA
 ; COUNTRY: USA
 ; ZIP: 98102
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: DOS
 ; SOFTWARE: FastSeq for Windows Version 2.0
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/10/124,089
 ; FILING DATE: 16-Apr-2002
 ; CLASSIFICATION: <Unknown>
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/811,481
 ; FILING DATE: <Unknown>
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Lingenfelter, Susan
 ; REGISTRATION NUMBER: P-41,156
 ; REFERENCE/DOCKET NUMBER: 95-36
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 206-442-6675
 ; TELEFAX: 206-442-6678
 ; TELEX: <Unknown>
 ; INFORMATION FOR SEQ ID NO: 34:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2328 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: cdna
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 34:
 ; US-10-124-089-34

Query Match 93.7%; Score 426.4; DB 16; Length 2328;
 Best Local Similarity 98.9%; Pred. No. 5.9e-100;
 Matches 449; Conservative 0; Mismatches 3; Indels 2; Gaps 2;
 Qy 1 GAAATCAGAGTTTAAATATGACACAAATTAATATATTTTGTATATCTCACACGGA-GNNT 59
 Db 2282 GAAATCAGAGTTTAAATATGACACAAATTAATATATTTTGTATATCTCACACGAGGTTT 2223
 Qy 60 CTCTTCAAACATAAGGAGTTAGAAATTAACAAGTAGGCATATGCTTCTATATTCAGATAA 119
 Db 2222 CTCTTCAAACATAAGGAGTTAGAAATTAACAAGTAGGCATATGCTTCTATATTCAGATAA 2163
 Qy 120 ATTCATTTTCGATTAAATTAATTCAGATAGAGAGAAATTAATTTTCGGAAGAAATGATA 179
 Db 2162 ATTCATTTTCGATTAAATTAATTCAGATAGAGAGAAATTAATTTTCGGAAGAAATGATA 2104
 Qy 180 GCTATATTAAGCAGATATTCATTAACAATACATGTAGACATAGCAATATTTTGGCA 239
 Db 2103 GCTATATTAAGCAGATATTCATTAACAATACATGTAGACATAGCAATATTTTGGCA 2044
 Qy 240 TCATTCTGCTCGCTCAGTAGGCGGTGTTCCCTCTGTTAGGCGCTTTGGAGAGTACCATCT 299
 Db 2043 TCATTCTGCTCGCTCAGTAGGCGGTGTTCCCTCTGTTAGGCGCTTTGGAGAGTACCATCT 1984
 Qy 300 ATCTAAGATGGAGGAATGCTGTGGAGAGGCGGGATGGAGGTGGCTTTTCTACGCTGAAC 359
 Db 1983 ATCTAAGATGGAGGAATGCTGTGGAGAGGCGGGATGGAGGTGGCTTTTCTACGCTGAAC 1924
 Qy 360 CCCACACAGGAAATCTGAGGCCACACAGCTGCTCTGCGCGCGCTTCCATGTGATCATC 419
 Db 1923 CCCACACAGGAAATCTGAGGCCACACAGCTGCTCTGCGCGCGCTTCCATGTGATCATC 1864
 Qy 420 CTGGTCAATGAAGTGAATGCTCTATTTCNGGGG 453
 Db 1863 CTGGTCAATGAAGTGAATGCTCTATTTCNGGGG 1830

RESULT 5
US-10-887-553A-435/c
; Sequence 435, Application US/10887553A
; Publication No. US20050085436A1
; GENERAL INFORMATION:
; APPLICANT: Garza, Dan
; APPLICANT: Li, Hao
; TITLE OF INVENTION: Method to treat conditions associated
; TITLE OF INVENTION: with insulin signalling dysregulation
; FILE REFERENCE: 4-33262
; CURRENT APPLICATION NUMBER: US/10/887,553A
; CURRENT FILING DATE: 2004-07-08
; PRIOR APPLICATION NUMBER: 60/485,883
; PRIOR FILING DATE: 2003-08-07
; NUMBER OF SEQ ID NOS: 1208
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 435
; LENGTH: 4767
; TYPE: DNA
; ORGANISM: human
US-10-887-553A-435

Query Match 93.6%; Score 426; DB 19; Length 4767;
Best Local Similarity 99.1%; Pred. No. 1e-99;
Matches 448; Conservative 0; Mismatches 2; Indels 2; Gaps 2;

QY	3	AATCAGAACTTTAATATGACACAAATTAATATATTTGTATATCTCACACCGGA-GNTTCT	61
DB	4758	AATCAGAACTTTAATATGACACAAATTAATATATTTGTATATCTCACACCGAGGTTTCT	4699
QY	62	CTTCAAACATAGGAGCTTGAATAATCAAGTAGGCATATGCTTCTATATTCAGATAAAT	121
DB	4698	CTTCAAACATAGGAGCTTGAATAATCAAGTAGGCATATGCTTCTATATTCAGATAAAT	4639
QY	122	TCATTTCGATTAATTAATATCCAGATAGAGAGAAATTTTCGAAAAAGAAATGATAGC	181
DB	4638	TCATTTCGATTAATTAATATCCAGATAGAGAGAAATTTTCGAAAAAGAAATGATAGC	4580
QY	182	TATATTAAGCAGATATTCATTACAATACGATGAGAGACATAAGCAATATTTGGCATC	241
DB	4579	TATATTAAGCAGATATTCATTACAATACGATGAGAGACATAAGCAATATTTGGCATC	4520
QY	242	ATTCTCTCGCTCAGTAGGCGGTTCCTCTGTTAGGGCCCTTGGAGAGTACCATCTAT	301
DB	4519	ATTCTCTCGCTCAGTAGGCGGTTCCTCTGTTAGGGCCCTTGGAGAGTACCATCTAT	4460
QY	302	CTAAGATGAGGAATGCTGTGGGAAGCGCGGATGGAGGTGCGTTTCTACGCTGAACCC	361
DB	4459	CTAAGATGAGGAATGCTGTGGGAAGCGCGGATGGAGGTGCGTTTCTACGCTGAACCC	4400
QY	362	CACACAGGAATCTGACGCGCCACACAGCTGCTCTCGCGCGCTTCCATGTATCATCT	421
DB	4399	CACACAGGAATCTGACGCGCCACACAGCTGCTCTCGCGCGCTTCCATGTATCATCT	4340
QY	422	GGTCAATGAAGTGAATTTGCTTATTCNGGGG	453
DB	4339	GGTCAATGAAGTGAATTTGCTTATTCNGGGG	4308

RESULT 6
US-10-060-036-1301/c
; Sequence 1301, Application US/10060036
; Publication No. US20030073144A1
; GENERAL INFORMATION:
; APPLICANT: Benson, Darin R.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Persing, David H.
; APPLICANT: Hepler, William T.
; APPLICANT: Jiang, Yugu
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY

; TITLE OF INVENTION: AND DIAGNOSIS OF PANCREATIC CANCER
; FILE REFERENCE: 210121.566
; CURRENT APPLICATION NUMBER: US/10/060,036
; CURRENT FILING DATE: 2002-01-30
; NUMBER OF SEQ ID NOS: 4560
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1301
; LENGTH: 282
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-060-036-1301

Query Match 56.0%; Score 254.8; DB 14; Length 282;
Best Local Similarity 98.2%; Pred. No. 6.7e-56;
Matches 278; Conservative 0; Mismatches 3; Indels 2; Gaps 2;

QY	11	GTTTAATATGACACAAATTAATATATTTGTATATCTCACACCGGA-GNTTCTCTTCAAAC	69
DB	282	GTTTAATATGACACAAATTAATATATTTGTATATCTCACACCGAGGTTTCTCTCAAAC	223
QY	70	ATAAGGAGTTAGAAATTAACAAGTAGGCATATGCTTCTATATTCAGATAAAATTCATTTCG	129
DB	222	ATAAGTAGTTAGAAATTAACAAGTAGGCATATGCTTCTATATTCAGATAAAATTCATTTCG	163
QY	130	ATTAATTAATATCCAGATAGAGAGAAATTTTCGAAAAAGAAATGATAGCTATATTA	189
DB	162	ATTAATTAATATCCAGATAGAGAGAAATTTTCGAAAAAGAAATGATAGCTATATTA	104
QY	190	ACGATATATTCATTACAATACCATGTAGAGACATAAGCAATATTTGGCATCATCTGTC	249
DB	103	ACGATATATTCATTACAATACCATGTAGAGACATAAGCAATATTTGGCATCATCTGTC	44
QY	250	CGCTCAGTAGGCGGTTCCTCTGTTAGGGCCCTTGGAGAGT	292
DB	43	CGCTCAGTAGGCGGTTCCTCTGTTAGGGCCCTTGGAGAGT	1

RESULT 7
US-10-106-698-2332/c
; Sequence 2332, Application US/10106698
; Publication No. US20030109690A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptides
; FILE REFERENCE: PA005P1
; CURRENT APPLICATION NUMBER: US/10/106,698
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: PCT/US00/26524
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/157,137
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: US 60/163,280
; PRIOR FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 8564
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 2332
; LENGTH: 455
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (8)..(8)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (86)..(86)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (93)..(93)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (95)..(96)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature

LOCATION: (99)..(99)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (126)..(126)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (148)..(148)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (180)..(180)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (225)..(225)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (375)..(375)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (387)..(387)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (425)..(425)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (438)..(438)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (454)..(454)
OTHER INFORMATION: n equals a,t,g, or c
US-10-106-698-2332

Query Match 14.2%; Score 64.4; DB 15; Length 455;
Best Local Similarity 90.5%; Pred. No. 1.8e-06;
Matches 76; Conservative 0; Mismatches 7; Indels 1; Gaps 1;
Qy 2 AAATCAGAGTTTATATATGACACAAATTAATATTTTGTATATCTCAGACCGGA-GNTTC 60
Db 147 AAATCAGAGTTTATATATGACACAAATTAATATTTTGTATATCTCAGACCGGA-GNTTC 88
Qy 61 TCTTCAACATAGAGGTTAGAAA 84
Db 87 TTTTCAACATAGAGGTTAGAAA 64

RESULT 8
US-10-240-485-93
Sequence 93, Application US/10240485
Publication No. US20030148327A1
GENERAL INFORMATION:
APPLICANT: OLEK, Alexander
APPLICANT: PIEPENBROCK, Christian
APPLICANT: BERLIN, Kurt
TITLE OF INVENTION: Diagnosis of Diseases Associated with
TITLE OF INVENTION: Metastasis
FILE REFERENCE: 5013.1007
CURRENT APPLICATION NUMBER: US/10/240,485
CURRENT FILING DATE: 2002-10-02
PRIOR APPLICATION NUMBER: PCT/EP01/03970
PRIOR FILING DATE: 2001-04-06
PRIOR APPLICATION NUMBER: DE 10019058.8
PRIOR FILING DATE: 2000-04-06
PRIOR APPLICATION NUMBER: DE 10019173.8
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: DE 10032529.7
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: DE 10043826.1
PRIOR FILING DATE: 2000-09-01
NUMBER OF SEQ ID NOS: 202
SEQ ID NO 93
LENGTH: 6811
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:

OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-240-485-93
Query Match 9.7%; Score 44.2; DB 15; Length 6811;
Best Local Similarity 51.5%; Pred. No. 1;
Matches 100; Conservative 0; Mismatches 94; Indels 0; Gaps 0;
Qy 11 GTTTAATATGACACAAATTAATATTTTGTATATCTCAGACCGGAGNTTCTCTTCAACA 70
Db 6368 GTTTAAAAAATAATATTTTAAATTTGATTTTTTATTGAGATTAATTTTGGATAGA 6427
Qy 71 TAAGGAGTTAGAAATTACAAGTAGGCATATGCTTCTATATTCAGATAAAATTCATTTCGA 130
Db 6428 GAAGTAGAAGAAATTTGAATTAGAATTAAGTTTTTTGTTAGTAAATTTTGTGT 6487
Qy 131 TTAATTAATTCAGATAGAGAGGTAATTTTCGAAAAAGAAATGATAGCTATATTTAAA 190
Db 6488 AAGAGTTATTTTAATGTGGAGAGGTTAATTTTTTACGAAATGATTTGTTTATTGTAA 6547
Qy 191 GCAGATATTCTATTA 204
Db 6548 AACGAGGTTTATTA 6561

RESULT 9

US-10-311-455-192/c
Sequence 192, Application US/10311455
Publication No. US20030143606A1
GENERAL INFORMATION:
APPLICANT: OLEK, Alexander
APPLICANT: PIEPENBROCK, Christian
APPLICANT: BERLIN, Kurt
TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
TITLE OF INVENTION: Cytosine methylation
FILE REFERENCE: 5013.1014
CURRENT APPLICATION NUMBER: US/10/311,455
CURRENT FILING DATE: 2002-12-16
PRIOR APPLICATION NUMBER: PCT/EP01/07537
PRIOR FILING DATE: 2001-07-02
PRIOR APPLICATION NUMBER: DE 10032529.7
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: DE 10043826.1
PRIOR FILING DATE: 2000-09-01
NUMBER OF SEQ ID NOS: 2424
SEQ ID NO 192
LENGTH: 11422
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-192

Query Match 9.7%; Score 44.2; DB 15; Length 11422;
Best Local Similarity 51.5%; Pred. No. 1.3;
Matches 100; Conservative 0; Mismatches 94; Indels 0; Gaps 0;
Qy 2 AAATCAGAGTTTAAATATGACACAAATTAATATTTTGTATATCTCAGACCGGAGNTTCT 61
Db 3770 AAATATATAATAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 3711
Qy 62 CTTCAACATAGAGGTTAGAAATTAAGTAGGCATATGCTTCTATATTCAGATAAAT 121
Db 3710 AATCATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 3651
Qy 122 TCATTTGATTAATTAATTCAGATAGAGAGGTAATTTTCGAAAAAGAAATCATAGC 181
Db 3650 CTCATTTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3591
Qy 182 TATATTAAGCAGA 195
Db 3590 AAACTACAAAATA 3577

```
RESULT 10
US-10-257-166-18/c
; Sequence 18, Application US/10257166
; Publication No. US20040023230A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Method and Nucleic Acids for Analysing the Methylation of
; TITLE OF INVENTION: Genes Implicated in Pharmacogenomics
; FILE REFERENCE: 5013.1011
; CURRENT APPLICATION NUMBER: US/10/257,166
; CURRENT FILING DATE: 2002-10-07
; PRIOR APPLICATION NUMBER: PCT/EP01/07470
; DE 10032529.7
; DE 10043826.1
; PRIOR FILING DATE: 2001-06-29
; 2000-06-30
; 2000-09-01
; NUMBER OF SEQ ID NOS: 178
; SEQ ID NO 18
; LENGTH: 11422
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-257-166-18

Query Match          9.7%; Score 44.2; DB 17; Length 11422;
Best Local Similarity 51.5%; Pred. No. 1.3;
Matches 100; Conservative 0; Mismatches 94; Indels 0; Gaps 0;

QY 2 AAATCAGAAGTTTAATATGACACAATTAATATATTTTGTATCTCCACACCGAGNTTCT 61
DB 3770 AAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATACTCT 3711

QY 62 CTTCAACATAGAGGAGTTAGAAATTACAAAGTAGGCATATGCTTCTCTATATTCAGATAAAAT 121
DB 3710 AATCAATTAATAAAATTTTAAATATAAATCAATATTTTATCTCTATTTTCATTTATTT 3651

QY 122 TCATTTTCGATTAATTAATTCAGATAGAGAGAGTAATTTTCGGAAGAAAGAAATGATAGC 181
DB 3650 CTCCTTCGATTAATCAATTAATACATAATAAACAATATCAAAATTTTAAAAATCTTTTAA 3591

QY 182 TATATTAAGCAGA 195
DB 3590 AAAACTACAAATA 3577

RESULT 11
US-10-433-793-27
; Sequence 27, Application US/10433793
; Publication No. US20040142334A1
; GENERAL INFORMATION:
; APPLICANT: Epigenomics AG
; TITLE OF INVENTION: Diagnose von mit Angiogenese assoziierten Krankheiten
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/433,793
; CURRENT FILING DATE: 2003-06-06
; NUMBER OF SEQ ID NOS: 212
; SEQ ID NO 27
; LENGTH: 37515
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-433-793-27

Query Match          9.7%; Score 44.2; DB 18; Length 37515;
Best Local Similarity 51.5%; Pred. No. 2.2;
Matches 100; Conservative 0; Mismatches 94; Indels 0; Gaps 0;

QY 11 GTTTAATATGACACAATTAATATATTTTGTATATCTCCACACCGAGNTTCTCTTCAACA 70
```

```
DB 6368 GTTTAAAAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 6427
QY 71 TAAGGAGTTAGAAATTAACAAGTAGGCATATGCTTCTCTATATTCAGATAAAATTCATTCCA 130
DB 6428 GAAGTAGAAAGAAATTTGAATTAGGATTAATAGTTTTTTGTTAGGTAATATTTTGTGT 6487
QY 131 TTAATTAATCCAGATAGAGAGAGTAATTTTCGAAAAAGAAATGATAGCTATATATAA 190
DB 6488 AAGAGTTATTTAAATGTTGGAGAGGTTAATTTTTCAGAAATGATATGTTTATTATTGTA 6547
QY 191 GCAGATATTCATTA 204
DB 6548 AACGAGGTTTATTA 6561

RESULT 12
US-10-027-632-180643
; Sequence 180643, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 180643
; LENGTH: 670
; TYPE: DNA
; ORGANISM: Human
; OTHER INFORMATION:
US-10-027-632-180643

Query Match          9.5%; Score 43.2; DB 13; Length 670;
Best Local Similarity 49.8%; Pred. No. 0.66;
Matches 108; Conservative 0; Mismatches 109; Indels 0; Gaps 0;

QY 9 AAGTTTAATATGACACAATTAATAATATTTTGTATATCTCACACCGAGNTTCTCTTCAAA 68
DB 283 AAAGAGAGAGTGGACATTTCAATTTATTTGGAAAAAATCAAGCTGGAATGGAATATCCAA 342
QY 69 CATAAGGAGTTAGAAATTAACAAGTAGGCATATGCTTCTCTATATTCAGATAAAATTCATTTC 128
DB 343 TAATACCAATTTTAAAAACAATAGTAAATCAAGACTTCTTCAGATAAAACAATGAATCA 402
QY 129 GATTAAATTAATTCAGATAGAGAGAGTAATTTTCGAAAAAGAAATGATAGCTATATTA 188
DB 403 CTACCATTAATAAGTAATTTTAAAAATAAATTTTGGAGAAAAAATAATATCCATTA 462
QY 189 AAGCAGATATTCATTACATACCATGATAGAGACATAA 225
DB 463 AAGAAGGCTGGAATGCAACAAGAAATACTGAGAAAA 499

RESULT 13
US-10-027-632-180643
; Sequence 180643, Application US/10027632
```

Publication No. US20030204075A9

GENERAL INFORMATION:

APPLICANT: Wang, David G.

TITLE OF INVENTION: Identification and Mapping of Single Nucleotide

POLYMORPHISMS IN THE HUMAN GENOME

FILE REFERENCE: 108827.129

CURRENT APPLICATION NUMBER: US/10/027,632

CURRENT FILING DATE: 2002-04-30

PRIOR APPLICATION NUMBER: US 60/218,006

PRIOR FILING DATE: 2000-07-12

PRIOR APPLICATION NUMBER: US 60/198,676

PRIOR FILING DATE: 2000-04-20

PRIOR APPLICATION NUMBER: US 60/193,483

PRIOR FILING DATE: 2000-03-29

PRIOR APPLICATION NUMBER: US 60/185,218

PRIOR FILING DATE: 2000-02-24

PRIOR APPLICATION NUMBER: US 60/167,363

PRIOR FILING DATE: 1999-11-23

PRIOR APPLICATION NUMBER: US 60/156,358

PRIOR FILING DATE: 1999-09-28

PRIOR APPLICATION NUMBER: US 60/146,002

PRIOR FILING DATE: 1999-08-09

NUMBER OF SEQ ID NOS: 325720

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 180643

LENGTH: 670

TYPE: DNA

ORGANISM: Human

US-10-027-632-180643

Query Match 9.5%; Score 43.2; DB 17; Length 670;
Best Local Similarity 49.8%; Pred. No. 0.66;
Matches 108; Conservative 0; Mismatches 109; Indels 0; Gaps 0;

Qy 9 AAGTTTATATGACACATTTAAATATATTTCTATATCTCACACCGGAGNTTCTCTCAA 68
Db 283 AAGAGGAGAGTGGACATTTCAAATTTTGGAAAAATCAAGCTTCTTCAGATAAACAACTGAATTCA 342
Qy 69 CATAGGAGTGTAGAAATACAAAGTAGGCATATGCTTCTATATTCAGATAAAATTCATTC 128
Db 343 TAATACCAATTTAAACAAATAGTAAATCAAGACTTCTTCAGATAAACAACTGAATTCA 402
Qy 129 GATTAATTAATTCAGATAGAGAGTAATTTTCGAAAGAAATGATAGCTATATTA 188
Db 403 CTACCATTAATAAGTAATTTAAAAAATAAAATTTGGGAGGAAAAAATAATCCATA 462
Qy 189 AAGCAGATATTCATACATACCATGTTAGACATAA 225
Db 463 AAGAGGTCTGGAATGCAACAAAGAAATACTGAGAAA 499

RESULT 14

US-09-925-298-217

Sequence 217, Application US/09925298

Publication No. US20020039764A1

GENERAL INFORMATION:

APPLICANT: Rosen et al.

TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies

FILE REFERENCE: PA103

CURRENT APPLICATION NUMBER: US/09/925,298

CURRENT FILING DATE: 2001-08-10

PRIOR APPLICATION NUMBER: PCT/US00/05881

PRIOR FILING DATE: 2000-03-08

PRIOR APPLICATION NUMBER: 60/124,270

PRIOR FILING DATE: 1999-03-12

NUMBER OF SEQ ID NOS: 846

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 217

LENGTH: 2815

TYPE: DNA

ORGANISM: Homo sapiens

US-09-925-298-217

Query Match 9.5%; Score 43; DB 9; Length 2815;

Best Local Similarity 56.4%; Pred. No. 1.4; Indels 0; Gaps 0;

Matches 79; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

Qy 24 CAATTAATAATATTTGTATATCTCACACCGGAGNTTCTCTTCAACATAAGGAGTTAGAA 83
Db 1957 CAATTTATGTATTTGAATATCAGCAAAATTTTCCATAATTTATCATTAATTTGTA 2016
Qy 84 ATTACAAGTAGGCATATGCTTCTTATATTCAGATAAAATTCATTTCCGATTAAATTC 143
Db 2017 ACCACATCCAGTGTCTACTCTTAGAGTTTCAGATGAATTTCTTAAAAATTAATA 2076
Qy 144 AGATAGAGAGAGTAATTTT 163
Db 2077 AACTCCATAGTACTAATTTT 2096

RESULT 15

US-10-102-806-217

Sequence 217, Application US/10102806

Publication No. US20030054421A1

GENERAL INFORMATION:

APPLICANT: Rosen et al.

TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies

FILE REFERENCE: PA103P1C1

CURRENT APPLICATION NUMBER: US/10/102,806

CURRENT FILING DATE: 2002-03-22

PRIOR APPLICATION NUMBER: 09/925,298

PRIOR FILING DATE: 2001-08-10

PRIOR APPLICATION NUMBER: PCT/US00/05881

PRIOR FILING DATE: 2000-03-08

PRIOR APPLICATION NUMBER: 60/124,270

PRIOR FILING DATE: 1999-03-12

NUMBER OF SEQ ID NOS: 846

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 217

LENGTH: 2815

TYPE: DNA

ORGANISM: Homo sapiens

US-10-102-806-217

Query Match 9.5%; Score 43; DB 14; Length 2815;

Best Local Similarity 56.4%; Pred. No. 1.4; Indels 0; Gaps 0;

Matches 79; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

Qy 24 CAATTAATAATATTTGTATATCTCACACCGGAGNTTCTTCAACATAAGGAGTTAGAA 83
Db 1957 CAATTTATGTATTTGAATATCAGCAAAATTTTCCATAATTTATCATTAATTTGTA 2016
Qy 84 ATTACAAGTAGGCATATGCTTCTTATATTCAGATAAAATTCATTTCCGATTAAATTC 143
Db 2017 ACCACATCCAGTGTCTACTCTTAGAGTTTCAGATGAATTTCTTAAAAATTAATA 2076
Qy 144 AGATAGAGAGAGTAATTTT 163
Db 2077 AACTCCATAGTACTAATTTT 2096

Search completed: May 5, 2005, 12:12:44

Job time : 635.302 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 4, 2005, 23:17:50 ; Search time 1925.91 Seconds
(without alignments)
8992.766 Million cell updates/sec

Title: US-08-731-499-8
Perfect score: 455
Sequence: 1 GAATCAGAGCTTAATATG.....ATTGCTCTATTTCGGGGT 455

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:
1: gb_est1:
2: gb_est2:
3: gb_hic:
4: gb_est3:
5: gb_est4:
6: gb_est5:
7: gb_est6:
8: gb_gss1:
9: gb_gss2:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	453	99.6	455	7	N78571 za93b11.sl
2	428	94.1	458	4	B1712448 ie08a03.x
3	428	94.1	471	1	AA661732 nv04b06.s
4	428	94.1	472	1	AI082095 oz38a05.s
5	428	94.1	478	2	BE044633 hq87e04.x
6	428	94.1	530	5	BU728730 UI-E-Cl1-
7	428	94.1	539	1	AI955290 wx32c06.x
8	428	94.1	557	5	BU069389 im21a05.x
9	428	94.1	563	1	AI828084 wk31f02.x
10	428	94.1	642	5	BQ019198 UI-H-DT1-
11	428	94.1	715	5	BM985176 UI-CF-EC1
12	428	94.1	2413	3	AL157451 Homo sapi
13	427	93.7	454	4	BM272161 ig39h05.x
14	426	93.7	499	6	CA775968 ip01h07.x
15	424.8	93.4	459	1	AI433898 tl17g11.x
16	424.8	93.4	461	4	BM353165 ig76e02.x
17	424.8	93.4	486	5	BQ778146 ll14f01.x
18	424.8	93.4	519	1	AI425012 tg50c03.x
19	424.8	93.4	561	5	BU077543 im38b12.x
20	424.8	93.4	586	7	CK822118 ig76e02.y
21	424.8	93.4	782	6	CB850986 UI-CF-EN1
22	420.2	92.4	551	5	BX104775 BX104775
23	417.2	91.7	449	1	AA563807 nj08h09.s
24	417	91.6	447	1	AA460668 zx64c11.s

C	25	416	91.4	458	5	BQ048938	AGENCOURT
	26	416	91.4	564	2	AW292950	UI-H-BW0-
	27	415.2	91.3	457	1	AI457762	tj49c10.x
	28	415	91.2	441	1	AI276830	ql67f01.x
	29	413	90.8	522	4	BM662818	UI-E-Cl1-
	30	410.6	90.2	559	5	BQ787950	il45e02.x
	31	410.6	90.2	1116	1	AL572501	AL572501
	32	407.6	89.6	694	1	AV721709	AV721709
	33	406.4	89.3	458	4	BG059906	naf45f08.
	34	405	89.0	417	1	AI082653	ox59a02.s
	35	404	88.8	456	1	AI148116	qb43e06.x
	36	402.6	88.5	452	5	BQ786989	il52e05.x
	37	402.2	88.4	477	1	AA101715	zk95h12.s
	38	398	87.5	427	2	AW662836	hl82c03.x
	39	398	87.5	436	1	AI378227	tc65c09.x
	40	398	87.5	458	1	AI148121	qb43e12.x
	41	396.8	87.2	1906	3	CR604034	full-leng
	42	395.8	87.0	1822	3	CR590102	full-leng
	43	391.6	86.1	1155	1	AL573342	AL573342
	44	389.8	85.7	1844	3	CR596761	full-leng
	45	388	85.3	431	1	AI972872	wr44f07.x

ALIGNMENTS

RESULT 1
N78571
LOCUS N78571 455 bp mRNA linear EST 29-MAR-1996
DEFINITION za93b11.sl Soares fetal_lung_NDH19W Homo sapiens cDNA clone
IMAGE:300093 3', mRNA sequence.
ACCESSION N78571
VERSION N78571.1 GI:1241272
KEYWORDS EST
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE
AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Travaakis, E., Waterston, R., Williamson, A., Wohldmann, P. and Wilson, R.
The WashU-Merck EST Project
Unpublished (1995)
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LNL ; contact the IMAGE Consortium (info@image.lnl.gov) for further information.
Seq primer: ml3 -40 forward
High quality sequence stop: 390.

FEATURES
Location/Qualifiers
1..455
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="GDB:1245017"
/db_xref="taxon:9606"
/clone="IMAGE:300093"
/dev_stage="19 weeks"
/lab_host="DH10B (ampicillin resistant)"
/clone_lib="Soares fetal lung NDH19W"
/note="Organ: lung; Vector: pT7T3D (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5'-TGTTACCACTGAATGGGCGCGCAATTTTTTTTTTTT-3'], double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT7T3 vector

(Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by Bento Soares and M.Fatima Bernaldo. This library was constructed from the same fetus as the fetal heart library, Soares fetal heart NBH19W."

ORIGIN

Query Match 99.6%; Score 453; DB 7; Length 455;
Best Local Similarity 100.0%; Pred. No. 1.5e-110;
Matches 455; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAAATCAGAAGTTTAAATATGACACAAATTAATTTGTATATCTCACACCGGAGNTTC 60
Db |||||
QY 1 GAAATCAGAAGTTTAAATATGACACAAATTAATTTGTATATCTCACACCGGAGNTTC 60
Db |||||

QY 61 TCTTCAACATAGGAGTTAGAAATACAGTAGGAGCATATGCTTCTATATTCAGATAAA 120
Db |||||

QY 61 TCTTCAACATAGGAGTTAGAAATACAGTAGGAGCATATGCTTCTATATTCAGATAAA 120
Db |||||

QY 121 TTCAATTCGATTAAATTAATTCAGATAGAGAGTAATTTTCGAAAAGAAATGATAG 180
Db |||||

QY 121 TTCAATTCGATTAAATTAATTCAGATAGAGAGTAATTTTCGAAAAGAAATGATAG 180
Db |||||

QY 181 CTATATTAAGCAGATATTCATTAACAATACCATGTAGACATACCAATATTTTGGCAT 240
Db |||||

QY 181 CTATATTAAGCAGATATTCATTAACAATACCATGTAGACATACCAATATTTTGGCAT 240
Db |||||

QY 241 CATTCGTCCGCTCAGTAGCGGTGTTCCCTCTGTGGAGGCGCTTTGGAGAGTACCATCTA 300
Db |||||

QY 241 CATTCGTCCGCTCAGTAGCGGTGTTCCCTCTGTGGAGGCGCTTTGGAGAGTACCATCTA 300
Db |||||

QY 301 TCTAAGATGGAGAAATGCTGTGGAGGCGGAGTGGAGTGGTCTTCTACGCTGAACC 360
Db |||||

QY 301 TCTAAGATGGAGAAATGCTGTGGAGGCGGAGTGGAGTGGTCTTCTACGCTGAACC 360
Db |||||

QY 361 CCACACAGAAATCTCGAGCCACACAGCTGCTCTGCGCGGCTTCCATGTGATCATCC 420
Db |||||

QY 361 CCACACAGAAATCTCGAGCCACACAGCTGCTCTGCGCGGCTTCCATGTGATCATCC 420
Db |||||

QY 421 TGGTCAATGAGTGAATTCCTATTTTCNGGGGT 455
Db |||||

QY 421 TGGTCAATGAGTGAATTCCTATTTTCNGGGGT 455
Db |||||

RESULT 2

BI712448 458 bp mRNA linear EST 11-MAR-2002
LOCUS 1e08a03.x1 HR85 islet Homo sapiens cDNA clone IMAGE:5086349 3',
mRNA sequence.
DEFINITION BI712448 GI:15688143
ACCESSION BI712448
VERSION
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE

AUTHORS Melton, D., Brown, J., Kenty, G., Permutt, A., Lee, C., Kaestner, K., Lemshke, I., Searce, M., Brestelli, J., Gradwohl, G., Clifton, S., Hillier, L., Marra, M., Pape, D., Wylie, T., Martin, J., Blistain, A., Schmitt, A., Theising, B., Ritter, E., Ronko, I., Bennett, J., Cardenas, M., Gibbons, M., McCann, R., Cole, R., Tsagareishvili, R., Williams, T., Jackson, Y. and Bowers, Y.
TITLE Endocrine Pancreas Consortium
JOURNAL Unpublished (2000)
COMMENT Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
Endocrine Pancreas Consortium
Harvard University, Howard Hughes Medical Institute
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge, MA 02138
Tel: 617-495-1812
Fax: 617-495-8557
Email: dmelton@biohp.harvard.edu

Library was constructed by Dr. Hiroshi Inoue DNA sequencing by: Washington University Genome Sequencing Center For information on obtaining a clone please contact: Dr. Hiroshi Inoue (hinoue@im.wustl.edu)

Seq primer: -40UP from Gibco

High quality sequence stop: 446.

FEATURES

Location/Qualifiers
1..458
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5086349"
/tissue_type="Purified pancreatic islet"
/lab_host="DH10B"
/clone_lib="HR85 islet"
/note="Organ: Pancreas; Vector: pBluescript SK(-); Site_1: NotI; Site_2: XhoI; cDNA made by oligo-dT priming. -1kb. 5' Size-selected on agarose gel. Average insert size ~1kb. 5' XhoI site was destroyed after directional cloning. Amplified once. Contact information: Hiroshi Inoue, MD, Metabolism Div. (Alan Permutt Lab), Washington University School of Medicine, Box 8127, 660 South Euclid Ave., St. Louis, MO 63110, E-mail: hinoue@imgate.wustl.edu, Tel: 314-362-1916, Fax: 314-747-2692."

ORIGIN

Query Match 94.1%; Score 428; DB 4; Length 458;
Best Local Similarity 99.1%; Pred. No. 7.6e-104;
Matches 450; Conservative 0; Mismatches 2; Indels 2; Gaps 2;

QY 1 GAAATCAGAAGTTTAAATATGACACAAATTAATTTGTATATCTCACACCGGA-GNTT 59
Db |||||

QY 4 GAAATCAGAAGTTTAAATATGACACAAATTAATTTGTATATCTCACACCGGAGTTT 63
Db |||||

QY 60 CTCITCAACATAAGAGTTAGAAATTAACAAGTAGCATATGCTTCTATATTCAGATAA 119
Db |||||

QY 64 CTCITCAACATAAGAGTTAGAAATTAACAAGTAGCATATGCTTCTATATTCAGATAA 123
Db |||||

QY 120 ATTCAATTCGATTAAATTAATTCAGATAGAGAGTAATTTTCGAAAAGAAATGATA 179
Db |||||

QY 124 ATTCAATTCGATTAAATTAATTCAGATAGAGAGTAATTTTCGAAAAGAAATGATA 182
Db |||||

QY 180 GCTATATTAAGCAGATATTCATTAACAATACCATGTAGAGACATAAGCAATATTTTGGCA 239
Db |||||

QY 183 GCTATATTAAGCAGATATTCATTAACAATACCATGTAGAGACATAAGCAATATTTTGGCA 242
Db |||||

QY 240 TCATTCGTCCGCTCAGTAGGCGGTTCCTCTCTGTTAGGCGCTTTGGAGAGTACCATCT 299
Db |||||

QY 243 TCATTCGTCCGCTCAGTAGGCGGTTCCTCTCTGTTAGGCGCTTTGGAGAGTACCATCT 302
Db |||||

QY 300 ATCTAAGATGGAGGAATGCTGTGGAGGCGGAGTGGAGTGGTTCCTACGCTGAAC 359
Db |||||

QY 303 ATCTAAGATGGAGGAATGCTGTGGAGGCGGAGTGGAGTGGTTCCTACGCTGAAC 362
Db |||||

QY 360 CCCACACAGGAATCTGCGCCACACAGCTGCTCTGCGCGGCTTCCATGTGATCATC 419
Db |||||

QY 363 CCCACACAGGAATCTGCGCCACACAGCTGCTCTGCGCGGCTTCCATGTGATCATC 422
Db |||||

QY 420 CTGGTCAATGAAGTGAATTCCTATTTTCNGGGG 453
Db |||||

QY 423 CTGGTCAATGAAGTGAATTCCTATTTTCNGGGG 456
Db |||||

RESULT 3

AA661732 471 bp mRNA linear EST 12-NOV-1997
LOCUS nv04b06.s1 NCI_CGAP_Pr22 Homo sapiens cDNA clone IMAGE:1219187 3',
mRNA sequence.
DEFINITION AA661732
ACCESSION AA661732
VERSION AA661732.1 GI:2615823
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 471)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation by: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 470.
Location/Qualifiers
1. .471
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clones="IMAGE:121918"
/sex="male"
/tissue_type="normal prostate"
/lab_host="DH10B"
/clone_lib="NCI CGAP Pr22"
/note="Organ: prostate; Vector: p7T73D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from normal prostate bulk tissue, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified p7T73 vector. Library is normalized, and was constructed by Bento Soares and M. Fatima Bonaldo."
ORIGIN
Query Match 94.1%; Score 428; DB 1; Length 471;
Best Local Similarity 99.1%; Pred. No. 7.6e-104;
Matches 450; Conservative 0; Mismatches 2; Indels 2; Gaps 2;
Qy 1 GAAATCAGAAGTTTAAATGACACAATTAATATATTTTGTATATCTCACACCGAGNTT- 59
Db 1 GAAATCAGAAGTTTAAATGACACAATTAATATATTTTGTATATCTCACACCGAGTTG 60
Qy 60 CTCTTCAACATAGGAGTTAGAAATACAAAGTAGGCATATGCTTCTTATATTCAGATAA 119
Db 61 CTCTTCAACATAGGAGTTAGAAATACAAAGTAGGCATATGCTTCTTATATTCAGATAA 120
Qy 120 ATTCATTTTCGATTAAATTAATTCAGATAGAGAGAGTAATTTTCGGAAAAAGAAATGATA 179
Db 121 ATTCATTTTCGATTAAATTAATTCAGATAGAGAGAGTAATTTTTCGGAAAAAGAAATGATA 179
Qy 180 GCTATATTAAGCAGATATTCATTACAAATACCATGTAGAGACATAAGCAATATTTTGGCA 239
Db 180 GCTATATTAAGCAGATATTCATTACAAATACCATGTAGAGACATAAGCAATATTTTGGCA 239
Qy 240 TCATTTCTCGCTCAGTAGCGCGTGTCCCTCTGTGTAGGCGCTTTGGAGAGTACCATCT 299
Db 240 TCATTTCTCGCTCAGTAGCGCGTGTCCCTCTGTGTAGGCGCTTTGGAGAGTACCATCT 299
Qy 300 ATCTAAGATGGAGGAATCTGTGGAGGCGGATGGAGTGGTGTTCACGCTGAAC 359
Db 300 ATCTAAGATGGAGGAATCTGTGGAGGCGGATGGAGTGGTGTTCACGCTGAAC 359
Qy 360 CCCACAGAGAAATCTGCAGGCCCAACACAGCTGCCTCTCGCGCGCGCTTCCATGTGATCATC 419
Db 360 CCCACAGAGAAATCTGCAGGCCCAACACAGCTGCCTCTCGCGCGCGCTTCCATGTGATCATC 419
Qy 420 CTGGTCAATGAAGTGAATTCCTATTTTCNGGGG 453
|||||

Db 420 CTGGTCAATGAAGTGAATTCCTATTTCTGGGG 453
RESULT 4
AI082095
LOCUS O238a05.s1 Soares_NhHMPu_S1 Homo sapiens cDNA clone IMAGE:1677584
DEFINITION 3', mRNA sequence.
ACCESSION AI082095
VERSION AI082095.1 GI:3418887
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 472)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Insert length: 562 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 464.
Location/Qualifiers
1. .472
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clones="IMAGE:1677584"
/tissue_type="Pooled human melanocyte, fetal heart, and pregnant uterus"
/lab_host="DH10B"
/clone_lib="Soares_NhHMPu_S1"
/note="Organ: mixed (see below); Vector: p7T73D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; Equal amounts of plasmid DNA from three normalized libraries (melanocyte 2NbHM, pregnant uterus NbHPU, and fetal heart NbH19W) were mixed, and ss circles were used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of I.M.A.G.E. clones 260232-265223, 340488-345479, and 484488-489479."
ORIGIN
Query Match 94.1%; Score 428; DB 1; Length 472;
Best Local Similarity 99.1%; Pred. No. 7.6e-104;
Matches 450; Conservative 0; Mismatches 2; Indels 2; Gaps 2;
Qy 1 GAAATCAGAAGTTTAAATGACACAATTAATATATTTTGTATATCTCACACCGAGNTT- 59
Db 4 GAAATCAGAAGTTTAAATGACACAATTAATATATTTTGTATATCTCACACCGAGTTGT 63
Qy 60 CTCTTCAACATAGGAGTTAGAAATACAAAGTAGGCATATGCTTCTTATATTCAGATAA 119
Db 64 CTCTTCAACATAGGAGTTAGAAATACAAAGTAGGCATATGCTTCTTATATTCAGATAA 123
Qy 120 ATTCATTTTCGATTAAATTAATTCAGATAGAGAGAGTAATTTTCGGAAAAAGAAATGATA 179
Db 124 ATTCATTTTCGATTAAATTAATTCAGATAGAGAGAGTAATTTTTCGGAAAAAGAAATGATA 182
Qy 180 GCTATATTAAGCAGATATTCATTACAAATACCATGTAGAGACATAAGCAATATTTTGGCA 239
Db 183 GCTATATTAAGCAGATATTCATTACAAATACCATGTAGAGACATAAGCAATATTTTGGCA 242
Qy 240 TCATTTCTCGCTCAGTAGCGCGTGTCCCTCTGTGTAGGCGCTTTGGAGAGTACCATCT 299
Db 243 TCATTTCTCGCTCAGTAGCGCGTGTCCCTCTGTGTAGGCGCTTTGGAGAGTACCATCT 302
|||||

```

QY 300 ATCTAAGATGGAGGAATGCTGTGGGAAGGCGGGATGGAGGTGCGCTTTTCTACGCTGAAC 359
DB 303 ATCTAAGATGGAGGAATGCTGTGGGAAGGCGGGATGGAGGTGCGCTTTTCTACGCTGAAC 362
QY 360 CCCACACAGGAAATCTGCAGCCACACACAGCTGCCCTCTGCGCGCCCTTCATGTGATCATC 419
DB 363 CCCACACAGGAAATCTGCAGCCACACACAGCTGCCCTCTGCGCGCCCTTCATGTGATCATC 422
QY 420 CTGGTCAATGAAGTGAATGCTCTATTTCTGCGG 453
DB 423 CTGGTCAATGAAGTGAATGCTCTATTTCTGCGG 456

RESULT 5
BE044633
LOCUS hq87e04.x1 NCI_CGAP_Thy3 478 bp mRNA linear EST 08-JUN-2000
DEFINITION mRNA sequence.
ACCESSION BE044633
VERSION BE044633.1 GI:8361686
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
JOURNAL National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
COMMENT Tumor Gene Index
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Klaus Kaserer, M.D., Chris Moskaluk, M.D.,
Ph.D., Michael R. Emmer-Buck, M.D., F.H.D.
CDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium/LLNL
Clone Distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL, send email to:
info@image.llnl.gov
Seq primer: -40UP from Gibco
High quality sequence stop: 417.

FEATURES
Location/Qualifiers
1..478
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3126366"
/tissue_type="follicular carcinoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NCI CGAP Thy3"
/note="Organ: thyroid; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.4 kb. Library constructed by Life
Technologies."

ORIGIN
Query Match 94.1%; Score 428; DB 2; Length 478;
Best Local Similarity 99.1%; Pred. No. 7.7e-104;
Matches 450; Conservative 0; Mismatches 2; Indels 2; Gaps 2;

QY 1 GAAATCAGAGTTTAATATGACACAAATTAATATATTTCTATATCTCACACCGGA-GNNTT 59
DB 14 GAAATCAGAGTTTAATATGACACAAATTAATATATTTCTATATCTCACACCGAGGTTT 73
QY 60 CTCTTCAACATAAGGAGTTAGAAAATTACAAGTAGGCATATGCTTCTATATTCAGATAA 119
DB 74 CTCTTCAACATAAGGAGTTAGAAAATTACAAGTAGGCATATGCTTCTATATTCAGATAA 133
QY 120 ATTCAATTCGATTAATAATTCAGATAGAGAGAAAGTAATTTTCGGAAAAAGAAATGATA 179
DB 134 ATTCAATTCGATTAATAATTCAGATAGAGAGAAAGTAATTTT-GGAAAAAGAAATGATA 192

```

```

QY 180 GCTATATTAAAGCAGATATTCAATACATACCATAGACATAGCAATATTTTGGCA 239
DB 193 GCTATATTAAAGCAGATATTCAATACATACCATAGACATAGCAATATTTTGGCA 252
QY 240 TCATTTCTCTCCGCTCAGTAGGCCGTGTTCCCTCTGTGTAGGCGCTTTGGAGAGTACATCT 299
DB 253 TCATTTCTCTCCGCTCAGTAGGCCGTGTTCCCTCTGTGTAGGCGCTTTGGAGAGTACATCT 312
QY 300 ATCTAAGATGGAGGAATGCTGTGGGAAGGCGGGATGGAGGTGCGCTTTTCTACGCTGAAC 359
DB 313 ATCTAAGATGGAGGAATGCTGTGGGAAGGCGGGATGGAGGTGCGCTTTTCTACGCTGAAC 372
QY 360 CCCACACAGGAAATCTGCAGCCACACACAGCTGCCCTCTGCGCGCCCTTCATGTGATCATC 419
DB 373 CCCACACAGGAAATCTGCAGCCACACACAGCTGCCCTCTGCGCGCCCTTCATGTGATCATC 432
QY 420 CTGGTCAATGAAGTGAATGCTCTATTTCTGCGG 453
DB 433 CTGGTCAATGAAGTGAATGCTCTATTTCTGCGG 466

RESULT 6
BE044633
LOCUS hq87e04.x1 NCI_CGAP_Thy3 478 bp mRNA linear EST 09-OCT-2002
DEFINITION UI-E-CL1-aep-b-05-0-UI.s1 UI-E-CL1 Homo sapiens cDNA clone
ACCESSION UI-E-CL1-aep-b-05-0-UI 3', mRNA sequence.
VERSION UI-E-CL1-aep-b-05-0-UI 3', mRNA sequence.
KEYWORDS EST.
SOURCE BU728730.1 GI:23650900
ORGANISM Homo sapiens (human)
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE 1 (bases 1 to 530)
JOURNAL Bonaldo,M.F., Lennon,G. and Soares,M.B.
MEDLINE Normalization and subtraction: two approaches to facilitate gene
PUBMED discovery
COMMENT Genome Res. 6 (9), 791-806 (1996)
97044477
889548
Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: bento-soares@uiowa.edu
Tissue Procurement: Dr. Gregg Hageman
CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research
Genetics (www.resgen.com).
Seq primer: M13 FORWARD
POLYA=Yes.

FEATURES
Location/Qualifiers
1..530
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UI-E-CL1-aep-b-05-0-UI"
/tissue_type="human retina"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/clone_lib="UI-E-CL1"
/note="Organ: eye; Vector: pT7T3-Pac (Pharmacia) with a
modified polylinker; Site 1: EcoR I; Site 2: Not I;
UI-E-CL1 is a normalized cDNA library containing the
following tissue(s): retina. The library was constructed
according to Bonaldo, Lennon and Soares, Genome Research,
6:791-806, 1996. First strand cDNA synthesis was primed
with an oligo-dT primer containing a Not I site. Double

```


stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT7T3-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)₁₈ tail. The sequence tag for this library is CGCGC. This library was created for the program, Gene Discovery in the Visual System, supported by National Eye Institute (NEI). TAG TISSUE=human retina TAG_LIB=UI-E-CL1 TAG_SEQ=CGCGC"

ORIGIN

Query Match	94.1%;	Score 428;	DB 5;	Length 530;
Best Local Similarity	99.1%;	Pred. No. 7.8e-104;		
Matches 450;	Conservative 0;	Mismatches 2;	Indels 2;	Gaps 2;

Qy	1	GAAATCAGAAGTTTAAATATGACACAAATTAATATATATTTGTATATCTCACACCGA-GNTT	59
Db	19	GAAATCAGAAGTTTAAATATGACACAAATTAATATATATTTGTATATCTCACACCGAGGTTT	78
Qy	60	CTCTTCAAAACATAAGGAGTTAGAAATTCACAAGTAGGCATATGCTTCCTATATTCAGATAA	119
Db	79	CTCTTCAAAACATAAGGAGTTAGAAATTCACAAGTAGGCATATGCTTCCTATATTCAGATAA	138
Qy	120	ATTCAATTCGATTAAATTTAAAATTCAGATAGAGAGAAGTAAATTTTCGGAAGAAAGTAATGATA	179
Db	139	ATTCATTTTCGATTAAATTTAAAATTCAGATAGAGAGAAGTAAATTTT-GGAAGAAAGTAATGATA	197
Qy	180	GCTATATTAAGCGAGATATTCATTACAATACCATGTAGAGACATAAGCAATATTTTGGCA	239
Db	198	GCTATATTAAGCGAGATATTCATTACAATACCATGTAGAGACATAAGCAATATTTTGGCA	257
Qy	240	TCATTCTGTCGGCTCAGTAGGCCGTGTTCCTCTCGGTAGGCGCTTTGGAGAGTAGCACTCT	299
Db	258	TCATTCTGTCGGCTCAGTAGGCCGTGTTCCTCTCGGTAGGCGCTTTGGAGAGTAGCACTCT	317
Qy	300	ATCTTAAGATGGAGAAATGCTGTGGGAAGGGCGGATGGAGGTGCGTTTTTCTPACGCTGAAC	359
Db	318	ATCTTAAGATGGAGAAATGCTGTGGGAAGGGCGGATGGAGGTGCGTTTTTCTPACGCTGAAC	377
Qy	360	CCACACAGGAAATCTCGAGGCCACACAGTGCCTCTCTGGCGCGCTTCCCATGTGATCATC	419
Db	378	CCACACAGGAAATCTCGAGGCCACACAGTGCCTCTCTGGCGCGCTTCCCATGTGATCATC	437
Qy	420	CTGTCTCAATGAAGTGAATTTGCTATTTCNGGGG	453
Db	438	CTGTCTCAATGAAGTGAATTTGCTATTTCNGGGG	471

RESULT 7	
ACCESSION	A1955290
LOCUS	wx32c06.x1 NCI_CGAP_Pit1 Homo sapiens cDNA clone IMAGE:2545354 3',
DEFINITION	linear mRNA 539 bp
ACCESSION	A1955290
VERSION	A1955290.1
KEYWORDS	GI:5747600
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS	1 (bases 1 to 539)
TITLE	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
JOURNAL	National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
COMMENT	Tumor Gene Index Unpublished (1997) Contact: Robert Strausberg, Ph.D. Email: cgapbs-x@mail.nih.gov Tissue Procurement: Chris Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., CDNA Library Preparation: Life Technologies, Inc. CDNA Library Arrayed by: Christa Prange, The I.M.A.G.E. Consortium DNA Sequencing by: Washington University Genome

```

Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bbrp/image/image.html
Insert length: 2000 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 414.
Location/Qualifiers
1. .539
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2545354"
/tissue_type="four pooled pituitary adenomas"
/lab_host="DH10B"
/clone_lib="NCI_CGAP_Pit1"
/note="Organ: brain; Vector: pCMV-SPORT6; Site_1: Salt;
Site 2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Library constructed by Life Technologies."

```

ORIGIN

Query Match	94.1%	Score 428;	DB 1;	Length 539;
Best Local Similarity	99.1%	Pred. No. 7.8e-104;		
Matches 450;	Conservative 0;	Mismatches 2;	Indels 2;	Gaps 2;
Qy	1	GAATCAGAGGTTTAAATATACACAAATTAATATATATTTGTATATCTCACACCGGA-GNTT	59	
Db	4	GAATCAGAGGTTTAAATATACACAAATTAATATATTTGTATATCTCACACCGGAGTTT	63	
Qy	60	CTCTTCAACACATAAGGAGTTTGAAGTTTACAAGTAGGCATATGCTTCTTATATTCAGATAA	119	
Db	64	CTCTTCAACACATAAGGAGTTTGAAGTTTACAAGTAGGCATATGCTTCTTATATTCAGATAA	123	
Qy	120	ATTCAITTCGATTAAATTTAAATTCAGATAGAGAGAAGTAAATTTTCGAAAAAGAAATGATA	179	
Db	124	ATTCAITTCGATTAAATTTAAATTCAGATAGAGAGAAGTAAATTTT-GGAAAAAGAAATGATA	182	
Qy	180	GCTATATTAAGCAGATATTTCAATTAACATACCATGTAGACACATAAGCAATATTTTGGCA	239	
Db	183	GCTATATTAAGCAGATATTTCAATTAACATACCATGTAGACACATAAGCAATATTTTGGCA	242	
Qy	240	TCATTTCTGTCGCTCAGTAGGCGGTGTTCCCTCTGTGTAGGCGCTTTGGAGAGTACCATCT	299	
Db	243	TCATTTCTGTCGCTCAGTAGGCGGTGTTCCCTCTGTGTAGGCGCTTTGGAGAGTACCATCT	302	
Qy	300	ATCTAAGATCGAGGAATGCTGTGGGAAGGCGGATGGAGGTGCGTTTTTCTACGCTGAAC	359	
Db	303	ATCTAAGATCGAGGAATGCTGTGGGAAGGCGGATGGAGGTGCGTTTTTCTACGCTGAAC	362	
Qy	360	CCCACACAGGAAATCTGTGAGCCCAACACAGTGCCTCTGTGCGCGGCTTCCATGTGATCATC	419	
Db	363	CCCACACAGGAAATCTGTGAGCCCAACACAGTGCCTCTGTGCGCGGCTTCCATGTGATCATC	422	
Qy	420	CTGGTCAATGAAGTAATTTGCTATTTCNGGG	453	
Db	423	CTGGTCAATGAAGTAATTTGCTATTTCNGGG	456	

RESULT 8	BU069389	557 bp	linear	EST 27-AUG-2002
LOCUS	im21a05.x1	Human insulinoma	Homo sapiens	cdna clone IMAGE:6035601
DEFINITION	3', mRNA sequence.			
ACCESSION	BU069389			
VERSION	BU069389.1	GI:22510578		
KEYWORDS	EST.			
SOURCE	Homo sapiens (human)			
ORGANISM	Homo sapiens			
	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
REFERENCE	1 (bases 1 to 557)			
AUTHORS	Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K., Lemishka,I., Scearc,M., Brestelli,J., Gradwohl,G., Clifton,S.,			

Hillier, L., Marra, M., Pape, D., Wylie, T., Martin, J., Blistain, A., Schmitt, A., Theising, B., Ritter, E., Ronko, I., Bennett, J., Cardenas, M., Gibbons, M., McCann, R., Cole, R., Tsagaris, R., Williams, T., Jackson, Y., and Bowers, Y.
Endocrine Pancreas Consortium
Unpublished (2000)

TITLE
JOURNAL
COMMENT

Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
Endocrine Pancreas Consortium
Harvard University, Howard Hughes Medical Institute
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge, MA 02138

Tel: 617-495-1812
Fax: 617-495-8557

Email: dmelton@biohp.harvard.edu

Library was constructed by Dr. J. Ferrer In vivo mass-excised to pBluescript SK- by Dr. H. Inoue DNA sequencing by: Washington University Genome Sequencing Center for information on obtaining a clone please contact: Dr. Hiroshi Inoue (hinoue@im.wustl.edu)

Seq primer: -40UP from Gibco

High quality sequence stop: 470.

FEATURES
source

1..557
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:603601"
/tissue_type="insulinoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="Human insulinoma"
/note="Organ: pancreas; Vector: pBluescript SK-; Site: 1: XhoI; Site 2: EcoRI; Constructed with lambda ZAPII system (Stratagene) by Dr. J. Ferrer, in vivo mass-excised to pBluescript SK- by Dr. H. Inoue following the Washington University protocol
(http://genome.wustl.edu/est/lambda_protocol.shtml).
Please contact Hiroshi Inoue, MD/PhD for further information on this library (Metabolism Division, Permutt Laboratory, Washington University School of Medicine, Box 8127, 660 S Euclid Ave, St. Louis, MO 63110). Note: this is a Washington University Pancreas EST project library."

ORIGIN

Query Match 94.1%; Score 428; DB 5; Length 557;
Best Local Similarity 99.1%; Pred. No. 7.9e-104;
Matches 450; Conservative 0; Mismatches 2; Indels 2; Gaps 2;

QY 1 GAAATCAGAGTTTAAATGACACAAATTAATATTTGTATATCTCACACCGGA-GNTT 59
DB 17 GAAATCAGAGTTTAAATGACACAAATTAATATTTGTATATCTCACACCGGAGTTT 76
QY 60 CTCCTCAACATAAGGAGTTAGAAATTAACAAGTAGGCATATGCTTCTATATTCAGATAA 119
DB 77 CTCCTCAACATAAGGAGTTAGAAATTAACAAGTAGGCATATGCTTCTATATTCAGATAA 136
QY 120 ATTCAATTCGATTAATTAATTCAGATAGAGAGTAATTTTCGAAAAAGAAATGATA 179
DB 137 ATTCATTCGATTAATTAATTCAGATAGAGAGTAATTTT-GGAAAAGAAATGATA 195
QY 180 GCTATATTAAGCAGATATTCATTAACAATCCATGTAGACATAAGCAATATTTGGCA 239
DB 196 GCTATATTAAGCAGATATTCATTAACAATCCATGTAGACATAAGCAATATTTGGCA 255
QY 240 TCATTCGTCCGCTCAGTAGGCGCTGCTCCCTCGTAGGCGCTTTGGAGAGTACCATCT 299
DB 256 TCATTCGTCCGCTCAGTAGGCGCTGCTCCCTCGTAGGCGCTTTGGAGAGTACCATCT 315
QY 300 ATCTAAGATGGAGGAATGCTGTGGGAAGGCGGGATGGAGGTGCGTTTTCTACGCTGAAC 359
DB 316 ATCTAAGATGGAGGAATGCTGTGGGAAGGCGGGATGGAGGTGCGTTTTCTACGCTGAAC 375
QY 360 CCCACACAGGAATCTGCAGCCACACACAGTGCCTCTGCGCGCCCTTCCATGTGATCATC 419
DB 376 CCCACACAGGAATCTGCAGCCACACACAGTGCCTCTGCGCGCCCTTCCATGTGATCATC 435

QY 420 CTGGTCAATGAAGTGAATTCCTATTCCTATTCNGGG 453
DB 436 CTGGTCAATGAAGTGAATTCCTATTCCTATTCNGGG 469

RESULT 9

AI828084

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Unpublished (1998)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs@mail.nih.gov

Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D., Ph.D.

CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.

CDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:

www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 1234 Std Error: 0.00

Seq primer: -40UP from Gibco

High quality sequence stop: 482.

Location/Qualifiers

1..563

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:2413947"

/tissue_type="anaplastic oligodendroglioma"

/lab_host="DH10B"

/clone_lib="NCI-CGAP_Brn25"

/note="Organ: brain; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGAGCGCGCATAGGTTTTTTTTTTTTTTTTTTTTT T 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library is normalized, and was constructed by Bento Soares and M. Fatima Bonaldo."

ORIGIN

Query Match 94.1%; Score 428; DB 1; Length 563;

Best Local Similarity 99.1%; Pred. No. 7.9e-104;

Matches 450; Conservative 0; Mismatches 2; Indels 2; Gaps 2;

QY 1 GAAATCAGAGTTTAAATGACACAAATTAATATTTGTATATCTCACACCGGA-GNTT 59

DB 3 GAAATCAGAGTTTAAATGACACAAATTAATATTTGTATATCTCACACCGGAGTTT 62

QY 60 CTCCTCAACATAAGGAGTTAGAAATTAACAAGTAGGCATATGCTTCTATATTCAGATAA 119

DB 63 CTCCTCAACATAAGGAGTTAGAAATTAACAAGTAGGCATATGCTTCTATATTCAGATAA 122

QY 120 ATTCAATTCGATTAATTAATTCAGATAGAGAGTAATTTTCGAAAAAGAAATGATA 179

DB 123 ATTCAATTCGATTAATTAATTCAGATAGAGAGTAATTTT-GGAAAAGAAATGATA 181

DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Researchers may obtain clones from Research
 Genetics (www.reagen.com) or from Open Biosystems
 (www.openbiosystems.com).
 Seq primer: M13 FORWARD
 POLYA=Yes.

FEATURES

source
 1. .715
 /location/Qualifiers
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="UI-CF-EC1-acc-k-02-0-UI"
 /issue_type="Lung"
 /lab_host="Adult and Fetal"
 /clone_lib="UI-CF-EC1"
 /note="Organ: Lung; Vector: p773-Pac (Pharmacia) with a
 modified polylinker; Site 1: EcoR I; Site 2: Not I;
 UI-CF-EC1 is a normalized cDNA library containing the
 following tissue(s): Normal lung from adult and from fetal
 day 64, day 87, week 19 and week 42. The library was
 constructed according to Bonaldo, Lennon and Soares,
 Genome Research, 6:791-806, 1996. First strand cDNA
 synthesis was primed with an oligo-dT primer containing a
 Not I site. Double stranded cDNA was ligated to an EcoR I
 adaptor, digested with Not I, and cloned directionally
 into p773-Pac vector. The oligonucleotide used to prime
 the synthesis of first-strand cDNA contains a library tag
 sequence that is located between the Not I site and the
 (dT)18 tail. The sequence tag for this library is
 AAGTGGCTTAC.
 TAG_TISSUE=Normal Lung Epithelial Cells Tissue nos 369-371
 and 380-383
 TAG_L1B=UI-CF-EC1
 TAG_SEQ=AAGTGGCTTAC"

ORIGIN

Query Match 94.1%; Score 428; DB 5; Length 715;
 Best Local Similarity 99.1%; Pred. No. 8.2e-104;
 Matches 450; Conservative 0; Mismatches 2; Indels 2; Gaps 2;
 QY 1 GAAATCAGAGTTTAAATGACACAAATTAATATTTCTATATCTCACACCGGA-GNNT 59
 Db 19 GAAATCAGAGTTTAAATGACACAAATTAATATTTCTATATCTCACACCGGAGTTT 78
 QY 60 CTCCTCAACATAGAGGTTAGAAATTACAAGTAGGCATATGCTTCCTATATTCAGATA 119
 Db 79 CTCCTCAACATAGAGGTTAGAAATTACAAGTAGGCATATGCTTCCTATATTCAGATA 138
 QY 120 ATTCTATTCGATTAAATTAATCCAGATAGAGAGTAATTTTCGAAAAGAAATGATA 179
 Db 139 ATTCTATTCGATTAAATTAATCCAGATAGAGAGTAATTTTCGAAAAGAAATGATA 197
 QY 180 GCTATATTAAGCAGATATTCATTACATACCATTAGACATAGCAATATTTTGGCA 239
 Db 198 GCTATATTAAGCAGATATTCATTACATACCATTAGACATAGCAATATTTTGGCA 257
 QY 240 TCATTCCTGCTCAGTAGGCGGTGTTCCCTCTGCTAGGCGCTTTGGAGAGTACCATT 299
 Db 258 TCATTCCTGCTCAGTAGGCGGTGTTCCCTCTGCTAGGCGCTTTGGAGAGTACCATT 317
 QY 300 ATCTAGATGGAGGAATGCTGTGGGAAGCGGGATGGAGGTGCGTTTCTACGCTGAAC 359
 Db 318 ATCTAGATGGAGGAATGCTGTGGGAAGCGGGATGGAGGTGCGTTTCTACGCTGAAC 377
 QY 360 CCCACACAGGAATCTGCAGCCACACACAGCTGCCTCTGCGCGCCCTTCCATGTGATCAT 419
 Db 378 CCCACACAGGAATCTGCAGCCACACACAGCTGCCTCTGCGCGCCCTTCCATGTGATCAT 437
 QY 420 CTGGTCAATGAAGTGAATTTGCTCTATTTTCNGGG 453
 Db 438 CTGGTCAATGAAGTGAATTTGCTCTATTTCTGGGG 471

RESULT 12
 HSM802434/c
 LOCUS 2413 bp mRNA linear HTC 22-SEP-2004
 DEFINITION Homo sapiens mRNA; cDNA DKFZp761A0712 (from clone DKFZp761A0712).
 ACCESSION AL157451
 VERSION AL157451.1 GI:7018465
 KEYWORDS HTC.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 2413)
 AUTHORS Ansong,W., Krieger,S., Regiert,T., Rittmueller,C., Schwager,B.,
 Mewes,H.W., Weil,B., Amid,C., Osanger,A., Fobo,G., Han,M. and
 Wiemann,S.
 The German cDNA Consortium
 Direct Submission
 Submitted (22-SEP-2004) MIPS, Ingolstaedter Landstr.1, D-85764
 Neuherberg, GERMANY
 Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
 Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
 Heidelberg/Germany) within the cDNA sequencing consortium of the
 German Genome Project.
 This clone (DKFZp761A0712) is available at the RZPD Deutsches
 Ressourcenzentrum fuer Genomforschung GmbH in Berlin, Germany.
 Please contact RZPD for ordering:
 http://www.rzpd.de/cgi-bin/products/cl.cgi?cloneID=DKFZp761A0712
 Further information about the clone and the sequencing project is
 available at http://mips.gsf.de/projects/cdna/.
 Location/Qualifiers
 1. .2413
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="RZPD:DKFZp761A0712"
 /db_xref="taxon:9606"
 /clone="DKFZp761A0712"
 /tissue_type="amygdala"
 /clone_lib="761 (synonym: hamy2). Vector pSport1; host
 DH10B; sites NotI + SalI"
 /dev_stage="adult"
 /note="protein tyrosine phosphatase, receptor type, N
 polypeptide 2, N-terminus truncated"
 1. .2413
 /gene="DKFZp761A0712"
 <12. .749
 /gene="DKFZp761A0712"
 /codon_start=1
 /product="hypothetical protein"
 /protein_id="CAB75665.2"
 /db_xref="GI:50978458"
 /db_xref="GOA:Q9NSR5"
 /db_xref="UniProt/TREMBL:Q9NSR5"
 /translation="VPKRSLAVLTVDHSRVLLKKAENSHSHSDYINAGPIMDHPNRP
 AVIATQGPLPATVADFMQWVSGCVIWMITPLAENGVRQCYHYWPDGSLYHIE
 VNLVEHLWCEDFLVRSFVNLQTNRTVTQFHLSDYDRGVSPSSRSLDPRKV
 NKCYGRGSCPIIVHCSDGSRGTYVLDIMVLNKAAGKAKEIDIAATLEHLDRPQM
 VOTKEQFEFALTAVAAEVNAILKALPQ"

ORIGIN

Query Match 94.1%; Score 428; DB 3; Length 2413;
 Best Local Similarity 99.1%; Pred. No. 1e-103;
 Matches 450; Conservative 0; Mismatches 2; Indels 2; Gaps 2;
 QY 1 GAAATCAGAGTTTAAATGACACAAATTAATATTTGTTATATCTCACACCGGA-GNNT 59
 Db 2397 GAAATCAGAGTTTAAATGACACAAATTAATATTTGTTATATCTCACACCGGAGTTT 2338
 QY 60 CTCCTCAACATAGAGGTTAGAAATTACAAGTAGGCATATGCTTCCTATATTCAGATA 119
 Db 2337 CTCCTCAACATAGAGGTTAGAAATTACAAGTAGGCATATGCTTCCTATATTCAGATA 2278

Qy 120 ATTCATTCGATTAAATTAATTCAGATAGAGAGAGTAATTTTCGGAAGAAATGATA 179
 Db 2277 ATTCATTCGATTAAATTAATTCAGATAGAGAGAGTAATTTTCGGAAGAAATGATA 2219
 Qy 180 GCTATATTAAGAGAGATATTCATTAACATACATGTAGAGACATAAGCAATATTTTGCA 239
 Db 2218 GCTATATTAAGAGAGATATTCATTAACATACATGTAGAGACATAAGCAATATTTTGCA 2159
 Qy 240 TCATTCCTCGCTCAGTAGGCGGTTCCTCTGTGTAGGCGCTTTGGAGAGTACCATCT 299
 Db 2158 TCATTCCTCGCTCAGTAGGCGGTTCCTCTGTGTAGGCGCTTTGGAGAGTACCATCT 2099
 Qy 300 ATCTAAGATGAGGAATGCTCTGGAGGCGGATGGAGGTGGTTCCTACCGCTGAAC 359
 Db 2098 ATCTAAGATGAGGAATGCTCTGGAGGCGGATGGAGGTGGTTCCTACCGCTGAAC 2039
 Qy 360 CCCACAGAGAAATCTGCAGGCCACACAGCTGCCTCTGCGCGGCTTCCATGTGATCATC 419
 Db 2038 CCCACAGAGAAATCTGCAGGCCACACAGCTGCCTCTGCGCGGCTTCCATGTGATCATC 1979
 Qy 420 CTGTCAATGAAGTGAATTCCTATTTTCNGGGG 453
 Db 1978 CTGTCAATGAAGTGAATTCCTATTTTCNGGGG 1945

RESULT 13
 BM272161
 LOCUS ig39h05.x1 HR85 islet Homo sapiens cDNA 3', mRNA linear EST 20-DEC-2001
 ACCESSION BM272161
 VERSION BM272161.1 GI:17965439
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 454)
 Lemisha,I., Scarce,M., Brestelli,J., Gradwohl,G., Clifton,S.,
 Hillier,L., Marra,M., Pape,D., Wyllie,T., Martin,J., Blistain,A.,
 Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J.,
 Cardenas,M., Gibbons,M., McCann,R., Cole,R., Tsagareishvili,R.,
 Williams,T., Jackson,Y. and Bowers,Y.
 Endocrine Pancreas Consortium
 Unpublished (2000)
 Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
 Endocrine Pancreas Consortium
 Harvard University, Howard Hughes Medical Institute
 Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
 MA 02138
 Tel: 617-495-1812
 Fax: 617-495-8557
 Email: dmelton@biohp.harvard.edu
 Library was constructed by Dr. Hiroshi Inoue DNA sequencing by:
 Washington University Genome Sequencing Center For information on
 obtaining a clone please contact: Dr. Hiroshi Inoue
 (hinoue@im.wustl.edu)
 Seq primer: -40UP from Gibco
 High quality sequence stop: 447.
 Location/Qualifiers
 1..454
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /tissue_type="Purified pancreatic islet"
 /lab_host="DH10B"
 /clone_lib="HR85 islet"
 /note="Organ: Pancreas; Vector: pBluescript SK(-); Site 1:
 Not1; Site 2: XhoI; cDNA made by oligo-dT priming.
 Size-selected on agarose gel. Average insert size ~1kb. 5'
 XhoI site was destroyed after directional cloning.
 Amplified once. Contact information: Hiroshi Inoue, MD,
 Metabolism Div. (Alan Permutt Lab), Washington University

FEATURES
source

1..454
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /tissue_type="Purified pancreatic islet"
 /lab_host="DH10B"
 /clone_lib="HR85 islet"
 /note="Organ: Pancreas; Vector: pBluescript SK(-); Site 1:
 Not1; Site 2: XhoI; cDNA made by oligo-dT priming.
 Size-selected on agarose gel. Average insert size ~1kb. 5'
 XhoI site was destroyed after directional cloning.
 Amplified once. Contact information: Hiroshi Inoue, MD,
 Metabolism Div. (Alan Permutt Lab), Washington University

School of Medicine, Box 8127, 660 South Euclid Ave., St.
 Louis, MO 63110, E-mail: hinoue@imgate.wustl.edu, Tel:
 314-362-1916, Fax: 314-747-2692."

ORIGIN

Query Match 93.8%; Score 427; DB 4; Length 454;
 Best Local Similarity 99.18; Pred. No. 1.4e-103;
 Matches 449; Conservative 0; Mismatches 2; Indels 2; Gaps 2;
 Qy 2 AAATCAGAAGTTTAATATGACACAATTAATAATATTTGTATATCTCACACCGGA-GNNTC 60
 Db 1 AAATCAGAAGTTTAATATGACACAATTAATAATATTTGTATATCTCACACCGGAGTTTC 60
 Qy 61 TCCTCAACATACAGGAGTTAGAAATTAACAAGTAGGCATATGCTTCCTATATTCAGATAAA 120
 Db 61 TCCTCAACATACAGGAGTTAGAAATTAACAAGTAGGCATATGCTTCCTATATTCAGATAAA 120
 Qy 121 TTCAATTCGATTAAATTAATTCAGATAGAGAGAAAGTAATTTTCGAAAAAGAAATGATAG 180
 Db 121 TTCAATTCGATTAAATTAATTCAGATAGAGAGAAAGTAATTTTCGAAAAAGAAATGATAG 179
 Qy 181 CTATATTAAGCAGATATTCATTAACAATACCATGTAGAGACATAAGCAATATTTTGGCAT 240
 Db 180 CTATATTAAGCAGATATTCATTAACAATACCATGTAGAGACATAAGCAATATTTTGGCAT 239
 Qy 241 CATTCCTCGCTCAGTAGGCGGTTCCTCTCTGCTAGGCGCTTTGGAGAGTACCATCTA 300
 Db 240 CATTCCTCGCTCAGTAGGCGGTTCCTCTCTGCTAGGCGCTTTGGAGAGTACCATCTA 299
 Qy 301 TCTAAGATGAGGAATGCTGTGGGAAGCGGGATGGAGGTGGTTCCTACGCTGAACC 360
 Db 300 TCTAAGATGAGGAATGCTGTGGGAAGCGGGATGGAGGTGGTTCCTACGCTGAACC 359
 Qy 361 CCACACAGGAAATCTGCAGGCCACACAGCTGCCTCTGCGCGGCTTCCATGTGATCATCC 420
 Db 360 CCACACAGGAAATCTGCAGGCCACACAGCTGCCTCTGCGCGGCTTCCATGTGATCATCC 419
 Qy 421 TGGTCAATGAAGTGAATTCCTATTTTCNGGGG 453
 Db 420 TGGTCAATGAAGTGAATTCCTATTTTCNGGGG 452

RESULT 14

CA775968 499 bp mRNA linear EST 03-DEC-2002
 CA775968
 LOCUS ip01h07.x1 HR85 islet Homo sapiens cDNA clone IMAGE:6134653 3',
 mRNA sequence.

CA775968
 ACCESSION CA775968.1 GI:26013512

VERSION CA775968

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 499)

REFERENCE

AUTHORS

Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K.,
 Lemisha,I., Scarce,M., Brestelli,J., Gradwohl,G., Clifton,S.,
 Hillier,L., Marra,M., Pape,D., Wyllie,T., Martin,J., Blistain,A.,
 Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J.,
 Cardenas,M., Gibbons,M., McCann,R., Cole,R., Tsagareishvili,R.,
 Williams,T., Jackson,Y. and Bowers,Y.
 Endocrine Pancreas Consortium
 Unpublished (2000)

TITLE

JOURNAL

COMMENT

Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
 Endocrine Pancreas Consortium
 Harvard University, Howard Hughes Medical Institute
 Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
 MA 02138
 Tel: 617-495-1812
 Fax: 617-495-8557
 Email: dmelton@biohp.harvard.edu
 Library was constructed by Dr. Hiroshi Inoue DNA sequencing by:
 Washington University Genome Sequencing Center For information on

obtaining a clone please contact: Dr. Hiroshi Inoue

(hinoue@im.wustl.edu)

Seq primer: -40UP from Gibco

High quality sequence stop: 455.

FEATURES

source
1. .499
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6134653"
/tissue_type="Purified pancreatic islet"
/lab_host="DH10B"
/clone_lib="HR85 islet"
/note="Organ: Pancreas; Vector: pBluescript SK(-); Site 1:
NotI; Site 2: XhoI; cDNA made by oligo-dT priming.
Size-selected on agarose gel. Average insert size -1kb. 5'
XhoI site was destroyed after directional cloning.
Amplified once. Contact information: Hiroshi Inoue, MD,
Metabolism Div. (Alan Permutt Lab), Washington University
School of Medicine, Box 8127, 660 South Euclid Ave., St.
Louis, MO 63110. E-mail: hinoue@imgate.wustl.edu, Tel:
314-362-1916, Fax: 314-747-2692."

ORIGIN

Query Match 93.7%; Score 426.4; DB 6; Length 499;
Best Local Similarity 98.9%; Pred. No. 2.1e-103;
Matches 449; Conservative 0; Mismatches 3; Indels 2; Gaps 2;
QY 1 GAAATCAGAAGTTTAAATGACACAAATTAATTTGTATATCTCACACCGGA-GNNT 59
Db 6 GAAATCAGAAGTTTAAATGACACAAATTAATTTGTATATCTCACACCGGAGTTT 65
QY 60 CTCCTCAACATAGAGAGTTAGAAATTAACAAGTAGGCATATGCTTCTTATATTCAGATAA 119
Db 66 CTCCTCAACATAGAGAGTTAGAAATTAACAAGTAGGCATATGCTTCTTATATTCAGATAA 125
QY 120 ATTCATTTCCGATTAATAATTCAGATAGAGAGTAAGTAATTTTCGAAAAGAAATGATA 179
Db 126 ATTCATTTCCGATTAATAATTCAGATAGAGAGTAAGTAATTTTCGAAAAGAAATGATA 184
QY 180 GCTATATTAAAGCAGATATTCATTAACAATACCATGTAGAGACATAAGCAATATTTTGGCA 239
Db 185 GCTATATTAAAGCAGATATTCATTAACAATACCATGTAGAGACATAAGCAATATTTTGGCA 244
QY 240 TCATTTCTGCGCTCAGTAGGCGGTGCTCCCTCTGCTAGGCGCTTTGGAGAGTACCATCT 299
Db 245 TCATTTCTGCGCTCAGTAGGCGGTGCTCCCTCTGCTAGGCGCTTTGGAGAGTACCATCT 304
QY 300 ATCTAAGATGGAGGAATGCTGTGGGAAGGCGGGATGGAGGTGCGTTTTCTACGCTGAAC 359
Db 305 ATCTAAGATGGAGGAATGCTGTGGGAAGGCGGGATGGAGGTGCGTTTTCTACGCTGAAC 364
QY 360 CCACACAGAAATCTGCAGCCACACACAGTGCCTCTGCGCGCCCTTCATGTGATCATC 419
Db 365 CCACACAGAAATCTGCAGCCACACACAGTGCCTCTGCGCGCCCTTCATGTGATCATC 424
QY 420 CTGGTCAATGAAGTGAATTTGCTATTTCNCGGG 453
Db 425 CTGGTCAATGAAGTGAATTTGCTATTTCNCGGG 458

RESULT 15

AI433898
LOCUS t117g11.x1 NCI_CGAP_Kid11 Homo sapiens cDNA clone IMAGE:2130788 3',
DEFINITION mRNA sequence.
ACCESSION AI433898
VERSION AI433898.1 GI:4292400
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 459)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapsb@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone Distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 2020 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 452.

FEATURES

source
1. .459
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2130788"
/lab_host="DH10B"
/clone_lib="NCI CGAP Kid11"
/note="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with
a modified polylinker; Site 1: Not I; Site 2: Eco RI;
Plasmid DNA from the normalized library NCI_CGAP Kid3 was
prepared, and ss circles were made in vitro. Following HAP
purification, this DNA was used as tracer in a subtractive
hybridization reaction. The driver was PCR-amplified cDNAs
from a pool of 5,000 clones made from the same library
(cloneIDs 1322376-1323911, 1456007-1456775, and
1500552-1502855). Subtraction by Bento Soares and M.
Fatima Bonaldo."

ORIGIN

Query Match 93.4%; Score 424.8; DB 1; Length 459;
Best Local Similarity 98.7%; Pred. No. 5.5e-103;
Matches 448; Conservative 0; Mismatches 4; Indels 2; Gaps 2;
QY 1 GAAATCAGAAGTTTAAATGACACAAATTAATTTGTATATCTCACACCGGA-GNNT 59
Db 3 GAAATCAGAAGTTTAAATGACACAAATTAATTTGTATATCTCACACCGGAGTTT 62
QY 60 CTCCTCAACATAGAGAGTTAGAAATTAACAAGTAGGCATATGCTTCTTATATTCAGATAA 119
Db 63 CTCCTCAACATAGAGAGTTAGAAATTAACAAGTAGGCATATGCTTCTTATATTCAGATAA 122
QY 120 ATTCATTTCCGATTAATAATTCAGATAGAGAGTAAGTAATTTTCGAAAAGAAATGATA 179
Db 123 ATTCATTTCCGATTAATAATTCAGATAGAGAGTAAGTAATTTTCGAAAAGAAATGATA 181
QY 180 GCTATATTAAAGCAGATATTCATTAACAATACCATGTAGAGACATAAGCAATATTTTGGCA 239
Db 182 GCTATATTAAAGCAGATATTCATTAACAATACCATGTAGAGACATAAGCAATATTTTGGCA 241
QY 240 TCATTTCTGCGCTCAGTAGGCGGTGCTCCCTCTGCTAGGCGCTTTGGAGAGTACCATCT 299
Db 242 TCATTTCTGCGCTCAGTAGGCGGTGCTCCCTCTGCTAGGCGCTTTGGAGAGTACCATCT 301
QY 300 ATCTAAGATGGAGGAATGCTGTGGGAAGGCGGGATGGAGGTGCGTTTTCTACGCTGAAC 359
Db 302 ATCTAAGATGGAGGAATGCTGTGGGAAGGCGGGATGGAGGTGCGTTTTCTACGCTGAAC 361
QY 360 CCACACAGAAATCTGCAGCCACACACAGCTGCCTCTGCGCGCCCTTCATGTGATCATC 419
Db 362 CCACACAGAAATCTGCAGCCACACACAGCTGCCTCTGCGCGCCCTTCATGTGATCATC 421
QY 420 CTGGTCAATGAAGTGAATTTGCTATTTCNCGGG 453
Db 422 CTGGTCAATGAAGTGAATTTGCTATTTCNCGGG 455

Search completed: May 5, 2005, 06:54:12
Job time : 1935.24 secs

This Page Blank (uspto)